TEL AVIV UNIVERSITY

The Iby and Aladar Fleischman Faculty of Engineering

The Zandman-Slaner School of Graduate Studies

ON-DEMAND TAXI SERVICE BEHAVIOR UPON A ROAD HIERARCHY

THE CASE OF DIDI IN CHENGDU, CHINA

A final project report submitted toward the degree of

Master of Science in Industrial Engineering

by

Gal Levy

This research was carried out in The Department of Industrial Engineering

Under the supervision of Dr. Mor Kaspi and Dr. David Epstein

February 2021

(25.2.2021)

Abstract

By analyzing a taxi rides dataset published by GAIA Initiative (DiDi's open data project), which includes both ride orders and trajectory data, this work tries to gain insights regarding the specific road hierarchy characteristics of Chengdu city, as it emerges from two of DiDi's core ride services (DiDi Express and DiDi Premier). The work relies on previous research done using the mentioned dataset, and expands it using data from known open source projects, such as Open Street Maps (OSM) and innovative dynamic visualization tools, which were not previously applied on the researched dataset.

Within its limitations, this work targets to achieve the following goals: 1) Understand the relationship between pick-up and drop-off locations and road hierarchy for intra-city ride-sharing trips. 2) Leverage data behavior findings to provide policy recommendation about ride-sharing services. As part of this work contribution, it first provides an algorithm to enrich pick-up and drop-off locations with their matching road type, in relation to the level of how major or minor it is in the city road hierarchy. In the core of this work, we define three hypotheses related to the relationships between pick-ups and drop-offs distributions in relation to the city road hierarchy, and show a significant tendency towards major roads in pick-ups, and minor roads in drop-offs, as the main contribution.

In addition, we choose to focus on shared rides as a strategic and emerging sub-group of this dataset, detect those using methods from previous research, and examine their specific behavior. Our results shows a significant change of the road hierarchy distribution between pick-ups and drop-offs of shared rides, although with trends which are yet to be determined. Finally, we analyze the unique temporal characteristics of rides in relation to the city road hierarchy, and the flow between different road types, in relation to rides departures and arrivals, revealing the relationship between intra-higher road levels and intra-lower road level, among other insights. We conclude this work by standing on the importance of the city road hierarchy and characteristics when designing ride-sharing services policies, and especially modern policies such as corner-to-corner and similar.

Table of Contents

List of Illustrations

List of Tables

1. Introduction

The cost of congestion in the United States alone is roughly $121 billion per year or 1% of GDP, which includes 5.5 billion hours of time lost to sitting in traffic and an extra 2.9 billion gallons of fuel burned. These estimates do not even consider the cost of other potential negative externalities such as the vehicular emissions, travel-time uncertainty, and a higher propensity for accidents. The large-scale adoption of smart phones and the decrease in cellular communication costs has led to the emergence of a new mode of urban mobility, namely mobility-on demand (MoD) systems, led by companies such as Uber, Lyft, and Via. These systems are able to provide users with a reliable mode of transportation that is catered to the individual and improves access to mobility to those who are unable to operate a personal vehicle, reducing the waiting times and stress associated with travel.

An emerging sub-category of MoD services is ride-sharing. Ride-sharing services are transforming urban mobility by providing timely and convenient transportation to anybody, anywhere and anytime. These services present enormous potential for positive societal impacts with respect to pollution, energy consumption, congestion, etc. Dynamic ride-share systems aim to bring together travelers with similar itineraries and time schedules on short-notice. These systems may provide significant societal and environmental benefits by reducing the number of cars used for personal travel and improving the utilization of available seat capacity. Effective and efficient optimization technology that matches drivers and riders in real-time is one of the necessary components for a successful dynamic ride-share system. Advanced services, such as Via or MOIA, offers corner-to-corner service, which means passengers won't get picked-up or dropped off at their exact location. Via, for example, reports that wait times for pick-ups are usually about 5 minutes, which is faster than the bus system. Depending on the user destination, he or she may need to walk that last block or two to reach their desired location.

DiDi is a China-based MoD app from the company DiDi Chuxing. It is now a global service operator and competes with the likes of Uber – it is sometimes called as the Chinese Uber. It has a stake in Bolt in Europe, it purchased Uber China in 2016 and the Brazilian 99 more recently in 2019. DiDi has become the one-stop app to go to for hailing cabs or private cars, with 30 million trips completed on DiDi's platform every day (more than 10 billion trips a year).

One significant hub of the "Chinese Uber" operates in Chengdu city, the capital of Sichuan province, located in southwest China. It has an area of 14,300 square kilometers and has a population of approximately 16.3 million, while urban population is evaluated in 11.2 million, as of 2019. There are over 5 million cars in Chengdu, more than any other city in China except Beijing (2020). The city relies heavily on public transportation including mobility on-demand services, with DiDi being the most popular service for ride-sourcing and ride-sharing, with more than 8.5 million users. It recently has been studied that the percentage of shared trips in the city can potentially be increased from 7.8% to 90.7%, and the percentage of time savings can reach 25.7% from 2.4%.

By analyzing a dataset published by GAIA Initiative (DiDi's open data project), which includes both orders and trajectory data, this work tries to gain insights regarding the specific road hierarchy characteristics of Chengdu city, as it emerges from two of DiDi's core ride services (DiDi Express and DiDi Premier). The work relies on previous research done using the mentioned dataset, and expands it using data from known open source projects, such as Open Street Maps (OSM) and innovative dynamic visualization tools, which were not previously applied on the researched dataset.

Within its limitations, this work targets to achieve the following goals: 1) Understand the relationship between pick-up and drop-off locations and road hierarchy for intra-city ride-sharing trips. 2) Leverage data behavior findings to provide policy recommendation about ride-sharing services. As part of this work contribution, it first provides an algorithm to enrich pick-up and drop-off locations with their matching road type, in relation to the level of how major or minor it is in the city road hierarchy. In the core of this work, we define three hypotheses related to the relationships between pick-ups and drop-offs distributions in relation to the city road hierarchy, and show a significant tendency towards major roads in pick-ups, and minor roads in drop-offs, as the main contribution.

In addition, we choose to focus on shared rides as a strategic and emerging sub-group of this dataset, detect those using methods from previous research, and examine their specific behavior. Our results shows a significant change of the road hierarchy distribution between pick-ups and drop-offs of shared rides, although with trends which are yet to be determined. Finally, we analyze the unique temporal characteristics of rides in relation to the city road hierarchy, and the flow between different road types, in relation to rides departures and arrivals, revealing the relationship between intra-higher road levels and intra-lower road level, among other insights. We conclude this work by standing on the importance of the city road hierarchy and characteristics when designing ride-sharing services policies, and especially modern policies such as corner-to-corner.

The rest of this paper is organized as follows: Section 2 provides a literature review of previous research done using the DiDi dataset, of mobility on-demand services and its characteristics, and of other data-driven research done in relation to city road hierarchy. Section 3 details the research goals of our work. The multi-step methodology process is presented in sec 4, which is followed by detailed presentation and analysis of the results section 5. The last section summarizes the contributions, limitations, and future research directions of the study.

2. Literature review

2.1 Data driven research related to DiDi Chuxing dataset

In order to find potential gaps for research we started by reviewing the previous work done using the DiDi Chuxing taxi dataset, which was published in late 2017 [22]. While a lot of data driven research uses different ridesourcing datasets, including some in collected in Chinese cities - only a few use this specific dataset, collected in the city of Chengdu. While all works performed some kind of spatiotemporal analysis, some focused to leverage it for urban clustering methods, others for structuring potential

substitution for public transit, but the majority chose to focus on ridesharing. This can be related to the fact that the mentioned dataset includes a small portion of shared rides, but we can also assume it is related to the emerging interest and potential ridesharing, which grew significantly in recent years. This part will review the highlights found on the mentioned DiDi Chuxing dataset.

Gao, Qingke, et al. (2019) [1] developed a clustering method to help discover the specific functions that exist within urban regions. This method applied the Gaussian Mixture Model (GMM) to classify regions' inflow and trip count characteristics. It regroups these urban regions using the Pearson Correlation Coefficient (PCC) clustering method based on those typical characteristics. Using 1 week of the DiDi Dataset (approximately 1.5M data points) they demonstrate that the method can discriminate between urban functional regions, by comparing the proportion of surface objects within each region. There are some innovations that arose from this experiment. First, it found the series curves of inflow and trip count are a better means to represent the spatiotemporal patterns of residential travel than using pick-ups and drop-offs. Second, it shows that the method flow of GMM and PCC could identify different regions effectively. Finally, it claims that Points of Interest (POIs) could be taken into consideration when defining a region's main function.

A recent paper by Kong et al. (2020) [2] develops a three-level structure to recognize the potential substitution or complementary effects of ridesourcing on public transit. This paper investigates the effects through exploratory spatiotemporal data analysis. The results show that 33.1% of DiDi trips have the potential to substitute for public transit. The substitution rate is higher during the day (8:00–18:00), and the trend follows changes in public transit coverage. The substitution effect is more exhibited in the city center and the areas covered by the subway, while the complementary effect is more exhibited in suburban areas as public transit has poor coverage. Further examination of the factors impacting the relationship indicates that housing price is positively associated with the substitution rate, and distance to the nearest subway station has a negative association with it, while the effects of most built environment factors become insignificant. Based on these findings, policy implications are drawn regarding the partnership between transit agencies and ridesourcing companies, the spatial-differentiated policies in the central and suburban areas.

On a different focus, a paper by Tu, Meiting, et al. (2019) [3] aims to explore the potential of ridesplitting during peak hours, using a ridesplitting trip identification algorithm based on a share-ability network developed. It evaluated the gap between the potential and actual scales of ridesplitting. The results show that the percentage of potential cost savings can reach 18.47% with an average delay of 4.76 min, whereas the actual percentage is 1.22% with an average delay of 9.86 min. The percentage of shared trips can be increased from 7.85% to 90.69%, and the percentage of time savings can reach 25.75% from 2.38%. This is the first investigation of the gap between the actual scale and the potential of ridesplitting on a city scale. The proposed ridesplitting algorithm also take passenger delays into consideration. Further this research argues that the quantitative benefits could encourage transportation management agencies and

transportation network companies to develop sensible policies to improve the existing ridesplitting services.

In their work, Li, Wenxiang, et al. (2019) [4] further aims to explore the characteristics and effects of ridesplitting using observed ridesourcing data provided by DiDi dataset. First, a ridesplitting trip identification (RTI) algorithm is developed to separate the shared rides from the single rides (non-ridesplitting orders) and understand the ridesplitting share from total rides and their durations. Second, a ridesplitting trajectory reconstruction (RTR) algorithm is proposed to estimate the ridesplitting effects on delays and detours. Furthermore, the article analyzes and compares the scales, spatiotemporal patterns and travel characteristics between shared rides and single rides. The results show that the current percentage of ridesplitting in ridesourcing is still low (6-7%), which may be explained by the extra delay (about 10 min on average), detour (about 1.55 km on average), and degraded travel time reliability caused by ridesplitting. In addition, built environment factors, such as density and development, are positively correlated with ridesplitting demand and delay, while the diversity factor (mixed land-use) is negatively correlated with both. The findings of this study help better understand the features of ridesplitting and develop strategies for improving its use in emerging ridesourcing services. Our work will rely on the mentioned RTI algorithm and further develop the characteristics between shared and single rides.

2.2 Data driven research related to city road type characteristics

In a study done by Zhang, Yingjia, et al. (2015) [12] the researchers used data obtained from 340 cities based on the Chinese OSM road network to explore OSM road geometry (road density) and road attributes (road type) and their relationship. In this paper, the Shannon–Wiener index was used to evaluate the diversity of road types, for which a total of 340 city units were included. This metric classified Chengdu city, as a provincial capital, as a first grade road type diversity, which strength the potential of our own work on Chengdu city, using OSM diverse road types. In the rest of this work we refer to road type mainly as road level.

A significant work that we draw inspiration from is a publication by Ravi Shenkar [5], which used a high-scale (1.3 billion records) NYC GPS originated taxi dataset, ranging from 2009 to 2016. He produced some interesting visualizations of pickup and drop-off locations. Within his major insights, he compared the pickups and drop-offs on a point by point basis, showing how the avenues in Manhattan have more taxi pickups than the cross streets, which have more drop-offs.

2.3 Mobility on Demand (MoD) and ride-sharing services

In the following part we will review the characteristics, terminology and importance of Mobility on Demand services, focusing on ride-sharing.

The cost of congestion in the United States alone is roughly $121 billion per year or 1% of GDP [7], which includes 5.5 billion hours of time lost to sitting in traffic and an extra 2.9 billion gallons of fuel burned. These estimates do not even consider the cost of other potential negative externalities such as the vehicular emissions [8], travel-time uncertainty [9], and a higher propensity for accidents [10]. The large-scale adoption of

smart phones and the decrease in cellular communication costs has led to the emergence of a new mode of urban mobility, namely mobility-on demand (MoD) systems, led by companies such as Uber, Lyft, and Via. These systems are able to provide users with a reliable mode of transportation that is catered to the individual and improves access to mobility to those who are unable to operate a personal vehicle, reducing the waiting times and stress associated with travel. One of the major inefficiencies of current MoD systems is their capacity limitation, typically restricted to two passengers. A recent study in New York City showed that up to 80% of the taxi trips in Manhattan could be shared by two riders, with an increase in the travel time of a few minutes.

In a previous review performed by Alonso-Mora, Javier, et al. (2016) [6], the importance of MoD is being presented. Ride-sharing services are transforming urban mobility by providing timely and convenient transportation to anybody, anywhere and anytime. These services present enormous potential for positive societal impacts with respect to pollution, energy consumption, congestion, etc. Dynamic ride-share systems aim to bring together travelers with similar itineraries and time schedules on short-notice. These systems may provide significant societal and environmental benefits by reducing the number of cars used for personal travel and improving the utilization of available seat capacity. Effective and efficient optimization technology that matches drivers and riders in real-time is one of the necessary components for a successful dynamic ride-share system.

Ride-sharing services can provide not only a very personalized mobility experience but also ensure efficiency and sustainability via large-scale ride pooling. Large-scale ride-sharing requires mathematical models and algorithms that can match large groups of riders to a fleet of shared vehicles in real time. The mentioned work results [6] showed significant results from a study performed in NYC (2,000 vehicles, which are 15% of the taxi fleet, of capacity 10 or 3,000 of capacity 4 can serve 98% of the demand within a mean waiting time of 2.8 min and mean trip delay of 3.5 min).

A study obtained from DiDi services as well [11], analyzed the environmental impacts of the increasingly popular ridesharing travel, by taking Beijing as the empirical context. In the mentioned study by Yu, Biying, et al. several findings are obtained, which can fill an important gap in the understanding of this emerging travel mode in megacities: 1) ridesharing is able to reduce energy consumption, CO2 emissions, and NOx emissions by 26.6 thousand tce, 46.2 thousand tons, and 235.7 tons, As for the indirect impacts related to industrial production due to users' potential attitude change towards purchasing new cars or replacing the old cars, substantial energy savings and emission reduction. 2) Ridesharing trips show very obvious regional and temporal characteristics, and are mainly contributed to undertake commuting trips. 3) Ridesharing service mainly help meet the demand for the mid- and long-length trips in Beijing, since the average service distance of ridesharing trips is about 17.7 km with more than 70% of trips is longer than 10 km. 4) Ridesharing has evident influence on passengers' current travel model choice and their future attitudes towards purchasing new vehicles or replacing the old vehicles.

2.4 Ride-sourcing Vs Ride-splitting services

There are some controversies and confusion concerning the differences between these on-demand ride services and ridesharing services. Ridesharing indicates that drivers are travelers who share similar origins/destinations with their riders for a common purpose of conserving resources, saving money, or saving time. On the other hand, ridesourcing is a for-hire commercial service that operates similarly to taxi services. There are many ridesourcing companies around the world, including Uber, Lyft, Grab, Ola and DiDi Chuxing. As of 2020, Uber was operating in 69 countries in over 900 metropolitan areas, with approximately 5 million drivers [24], while Lyft was operating in 644 cities in the US, with over 2 million drivers (2019) [25]. In the past nearly six years, DiDi Chuxing has served more than 450 million users with a full range of mobility services across 400 cities in China, including Taxi, Express, Premier, Hitch, Bus, Minibus, Designated Driving, Car Rental and Enterprise Solutions. Most ridesourcing companies have launched services that enable riders to share the ride and split the cost with other people taking a similar route. These relatively new services are called ridesplitting, "a form of ridesourcing where riders with similar origins and destinations are matched to the same ridesourcing driver and vehicle in real time" (Shaheen et al., 2016). Early examples of ridesplitting are Lyft Line and UberPool, which allow unrelated passengers with overlapping routes to split rides and fares (Shaheen et al., 2015). Riders pay about half the normal price in exchange for sharing the car with other riders and accepting a detour of a few minutes to pick up and drop off a second passenger or group of passengers (Sperling, 2018).

2.5 Corner-to-corner ride-sharing services

A few services emerged recently offer a distinction from other known services. Via operates similarly to Uber and Lyft, but with a few distinctive differences. The most obvious distinction is that it is set up as an actual ride-sharing service. Passengers, in most cases, will get in a car with strangers who are heading in the same direction as them. It offers corner-to-corner service, so they won't get dropped off at their exact location. Via also says wait times are usually about 5 minutes, which is faster than the bus system. Depending on your destination, you may need to walk that last block or two to reach your desired location [26]. MOIA, a service provided by Volkswagen mobility group which operates in 2 cities in Germany, provides corner-to-corner service as well, which in that case means that pick-up and drop-off are up to 250 meters far from the addresses specified by the user [27]. Gilibert, Mireia, et al published a recent case study using data collected from MOIA activity in Hanover [28].

As mentioned at the head of this review, while all previous work on the DiDi dataset performed some kind of spatiotemporal analysis, some focused to leverage it for urban clustering methods, others for structuring potential substitution for public transit, but the majority chose to focus on ridesharing. Out of the few on this sub-category, while some analyzed some of urban characteristics of Chengdu city, none of them noticed a significant characteristic, which is the hierarchy of the city road levels – a key aspect while designing advanced ridesharing driving policies, such as corner-to-corner and similar. By acknowledging the importance of this feature, and by combining graphical and statistical tools we try to define two research goals.

3. Research goals

Following the above we define the targets we aim to achieve in this work:

1 Understand taxi intra city trips pick-ups and drop-offs points' behavior with relations to road hierarchy and ride-sharing.
2 Leverage data behavior findings to provide policy recommendation about ride-sharing services.

4. Methodology

4.1 Case study - Chengdu city

The study area for this research locates is the Chinese city of Chengdu. As the capital of Sichuan province, Chengdu is located in southwest China. It has an area of 14,300 square kilometers and has a population of approximately 16.3 million, while urban population is evaluated in 11.2 million, as of 2019. In addition, Chengdu contains many ethnic groups and has residents from 55 ethnic minority groups. It comprises 11 administrative districts, 5 county-level cities and 5 counties. Chengdu is a commercial logistics center and a comprehensive transportation hub. Its gross domestic product (GDP) exceeded 1.7 billion yuan in 2020. There are over 5 million cars in Chengdu, more than any other city in China except Beijing (2020). Because citizens mainly travel within the Fourth Ring Road area in Chengdu, we selected the radius of 10K from city center as the study area for this research.

Figure 1: City center of Chengu city
http://www.chinatraveldiscovery.com/china-map/chengdu-map.htm

4.2 Data description

The data used in this study are from the DiDi GAIA Initiative [22], DiDi's open data project (DiDi Chuxing, 2017). The project shares the complete ride trajectory and order data of DiDi Express and DiDi Premier, two of DiDi Chuxing's core ridesourcing services, in the city of Chengdu, China, from November 1-30, 2016. The trajectory dataset contains fields such as anonymous driver ID, order ID, timestamp, longitude, and latitude, with an average sampling interval of 3 seconds (some sample data are shown in Table 1). The order dataset includes fields such as order ID, start and end timestamps, pick-up and drop-off locations (some sample data are shown in Table 2). Taking into considerations of its time range, area coverage and size, which contains 7,065,907 records, we can assume this dataset constitutes a significant sample of the population from throughout Chengdu (yet not necessarily a representative one).

Table 1: Trajectory sample data, DiDi Chuxing dataset

Field	Type	Sample	Comment
DriverID	String	glox.jrrlltBMvCh8nxqktdr2dtopmlH	Anonymized
OrderID	String	jkkt8kxniovIFuns9qrrlvst@iqnpkwz	Anonymized
Timestamp	String	1,501,584,540	Unix Timestamp, in seconds
Longitude	String	104.04392	GCJ-02 Coordinate System
Latitude	String	104.04392	GCJ-02 Coordinate System

Table 2: Orders sample data, DiDi Chuxing dataset. Fields marked in [*] were extracted during pre-processing on top of original data fields

Field	Type	Sample	Comment
OrderID	String	mjiwdgkqmonDFvCk3ntBpron5mwfrqvI	Anonymized
Ride_Start_Timestamp	String	1,501,581,031	Unix Timestamp (seconds)
Ride_End_Timestamp	String	1,501,582,195	Unix Timestamp (seconds)
Pick-up_Longitude	String	104.11225	GCJ-02 Coordinate System
Pick-up_Latitude	String	30.66703	GCJ-02 Coordinate System
Drop-off_Longitude	String	104.07403	GCJ-02 Coordinate System
Drop-off_Latitude	String	30.6863	GCJ-02 Coordinate System
Day_of_week*	Integer	4	[0,1,2..6] Range, Shifted 6AM-6AM
Pick-up_hour*	Integer	13	[0,1,2..23] Range
Drop-off_hour*	Integer	14	[0,1,2..23] Range
Duration*	Integer	35	In minutes
Air_Distance*	Float	3.54	KM, Euclidean distance

4.3 Data preparation

4.3.1 Data filtering

A total share of 13.6% filtered records were filtered out from original order dataset due to duplicate rides (identical order id, start time, end time, pick-up and drop-off coordinates). The remaining dataset size is 6,105,003 records (86.4% of original data).

4.3.2 GCJ-02 coordinate system and data offset

The GCJ-02 coordinate system is a geodetic datum formulated by the Chinese State Bureau of Surveying and Mapping, and based on WGS-84. It uses an obfuscation algorithm, which adds apparently random offsets to both the latitude and longitude, with the alleged goal of improving national security. There is a license fee associated with using this mandatory algorithm in China. A marker with GCJ-02 coordinates will be displayed at the correct location on a GCJ-02 map. However, the offsets can result in a 100 - 700 meter error from the actual location if a WGS-84 marker (such as a GPS location) is placed on a GCJ-02 map, or vice versa. [13]

4.3.3 Reverse transformation

GCJ-02 appears to use multiple high-frequency noises, effectively generating a transcendental equation and thus eliminating analytical solutions. However, the open-source "reverse" transformations [14] make use of the properties of GCJ-02 that the transformed coordinates are not too far from WGS-84 and are mostly monotonic related to corresponding WGS-84 coordinates. The rough method is reported to give some 1~2 meter accuracy for wgs2gcj. [13]

4.3.4 Shared rides extraction

4.3.4.1 DiDi ExpressPool

As mentioned above, the dataset is originated in 2 of DiDi's services: Express and Premier. DiDi express offers the option for a passenger to choose whether to share a ride with other passengers when s/he requests the ride. If the passenger chooses the ridesplitting option (also called ExpressPool), he will receive a discount of up to 40% on the price based on the likelihood of matching with another rider, regardless of whether the trips are successfully matched eventually. The payments to ridesplitting drivers are mostly based on the actual travel time and distance of the trip, with some additional rewards. By April 2018, the daily number of rides on DiDi ExpressPool surpassed 2.4 million in 60 cities (compared with the 25 million daily orders on all its services), which makes DiDi ExpressPool the main ridesplitting service in China. DiDi Premier, the second type of services we can find in the dataset, is considered to deliver a high-end, luxury rides with specially trained drivers.

4.3.4.2 Ridesplitting Trip Identification (RTI)

In order to explore the relation of ridesplitting to our research questions we would first need to detect the ridesplitting rides from the raw dataset. This would allow us to ask questions regarding shared rides versus single rides, as previously done on Li, Wenxiang, et al. [4]. For the full methodology, notations and pseudo-code we can refer to the sec 4.1, 4.2 in the original article [4], but we will mention the main definitions:

- **Single ride:** from a rider's perspective, the rider travels and bears the ride cost alone.
- **Shared ride:** from a rider's perspective, the rider shares the ride and splits the cost with other riders taking a similar route, as arranged by the application in real-time.
- **Ridesplitting trip:** from a driver's perspective, the driver receives multiple ride requests on a trip for which riders have similar origins or destinations; a ridesplitting trip consists of two or more shared rides starting from the time when the driver picks up the first customer to the time when the driver drops off the last customer (i.e. the vehicle becomes vacant again)

Our results after applying the RTI was almost identical to the original researchers' results, detecting 375,632 out of 375,661 original results (99.99%). Distribution of remaining shared ride characteristics were almost identical as well. Each shared ride detected was labeled as 'Shared' while others were labeled as 'Single'.

4.3.5 Open Street Maps

The OpenStreetMap (OSM) project is perhaps one of the most successful examples of crowdsourcing in the spatial domain. Streets comprise the single most important feature in the OSM database. Street network information is fundamental for many applications such as navigation, network analysis, and map generalization, just to name a few. The basic semantic information of a street is its class such as a motorway, a primary road, or a residential road, etc. This information indicates several things about a street: its possible neighborhoods, its permissible driving speeds, and the level of map generalization at which it should be displayed.

As shown in Table 3, the road network dataset was extracted using the *osmnx* library [17], which lists ~31K edges, after bounding the requested radius in 10K from city center and filtering only driving network. A complete data description is detailed in table 3. The driving network in the city of Chengdu includes 8 road levels ranging from highest (Motorway) to lowest (Living Street), and does not include paths, street and special roads which are not accessible for driving, such as service roads, pedestrian and others. For the purpose of binary classification, in some parts of the rest of this paper we will refer to the highest five types (Motorway to Tertiary) as "major roads" and the lower three (Unclassified to Living Street) as "minor roads". A detailed list of the driving road list can be found in table 4.

Table 3: Road network data of OSM

Field	Type	Sample	Comment
osmid	Int	99989683	Unique network edge ID
u	Int	359203175	OSM "From" node ID
v	Int	359203168	OSM "To" node ID
name	String	Shawan Road	Name of the road
geometry	String	LINESTRING (-21.93067 64.05665, -21.93067 64.0..)	(u,v) Location Info (GPS)
oneway	Binary	FALSE	Is this road one way
lanes	Integer	2	Number of lanes
highway	String	Primary	Road hierarchy level
length	Float	237.337	Total length (KM)
maxspeed	Integer	70	Max speed allowed

Table 4: Road hierarchy levels of OSM [15]

Road level	Description
Motorway	A restricted access major divided highway, normally with 2 or more running lanes plus emergency hard shoulder. Equivalent to the Freeway, Autobahn, etc.
Trunk	High performance or high importance roads that don't meet the requirement for motorway
Primary	A major highway linking large towns, in developed countries normally with 2 lanes. In areas with worse infrastructure road quality may be far worse. The traffic for both directions is usually not separated by a central barrier.
Secondary	A highway which is not part of a major route, but nevertheless forming a link in the national route network. In developed countries it normally has 2 lanes and the traffic for both directions is usually separated by a central line on the road.
Tertiary	Within larger urban settlements such as large towns or cities, tertiary roads link local centers of activity such as shops, schools, or suburbs. Low to moderate traffic.
Unclassified	Minor public roads typically at the lowest level of the interconnecting grid network. Used for roads used for local traffic, and for roads used to connect other towns, villages or hamlets.
Residential	roads that are used for accessing residential areas and in residential areas but which are not normally used as through routes
Living Street	A street where pedestrians have priority over cars, children can play on the street, maximum speed is low.

4.3.5.1 OSM data pre-processing

It was found that 204 out of 31K (0.6%) edges detected within the research area of Chengdu with no distinct "highway" label. Some examples can be found such as ['living_street', 'tertiary', 'residential'], ['motorway', 'trunk'] or just 'road'. These edges length are accounted for 1.5% of total research area road length and ~1% of rides were matched to these roads as nearest. Since there is no method of knowing the actual sub-

roads for these edges and their hierarchy level, two method were applied to handle these anomalies: 1) Split their length equally for each label in the list of multiple labels. 2) For all rides matched with a multi-list choose randomly (uniform distribution) a label from each multi-list and assign this label to the ride. This methods aim to minimize the error caused by the ambiguity of the indistinct labels. Based on this, and the relatively low share of these anomalies the integrity of the data should remain. Complete distribution of city road levels length following these steps can be shown in figure 2.

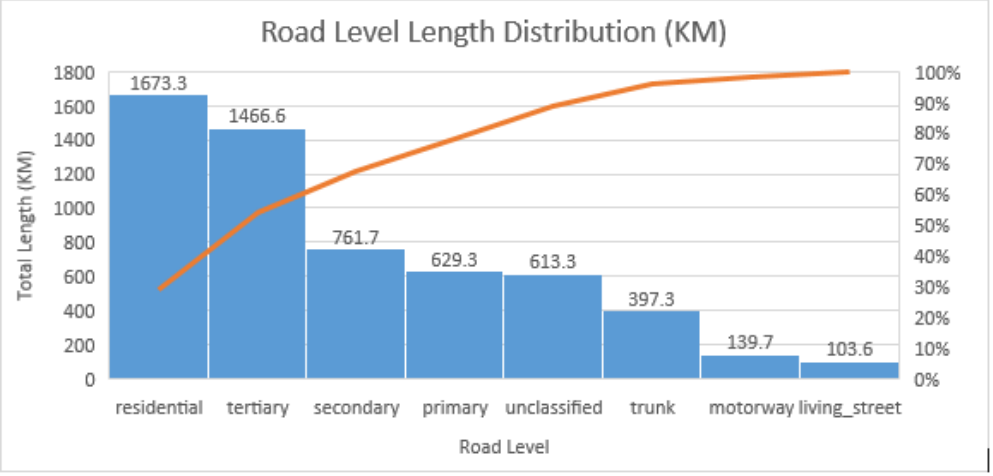

Figure 2: Road level Length (KM) distribution in Chengdu city, 10KM radios from city center

4.4 Exploratory data analysis using dynamic visualization

First, we would like to get a better understanding of our data and explore it. As mentioned in our literature review, the DiDi rides dataset was researched before by a few studies, and provided exploratory analysis on the original dataset. Statistical insights such as trajectory and order data feature analysis were provided by Li, Wenxiang, et al. [4], which also explored the characteristics of shared rides detected. A graphical representation of the pickups and drop-offs was created by Kong et al. [2], but none created a rich and dynamic visualization as the one we present on this dataset, at least for our knowledge.

4.4.1 Spatial analysis of DiDi pick-ups and drop-offs

We begin by plotting a pick-ups and drop-offs of 5.2 million rides in our filtered research area radius. Since the human eyes are incapable of absorbing this amount of information in a simple plot, we leverage the *datashader* [18] library. Each pixel on the display corresponds to certain histogram boundaries in the data. The library counts the number of data points that fall within those boundaries for each pixel, and this number is used to color the intensity of the pixel. From the results comparing pick-ups and drop-offs (Figure 3) we can spot a higher intensity around major roads in pick-ups, a pattern that emphasizes the city center road outline. In contrary, drop-offs plot seems to

portraits a more uniform distribution between major and minor roads, possibly more common in residential areas than pick-ups.

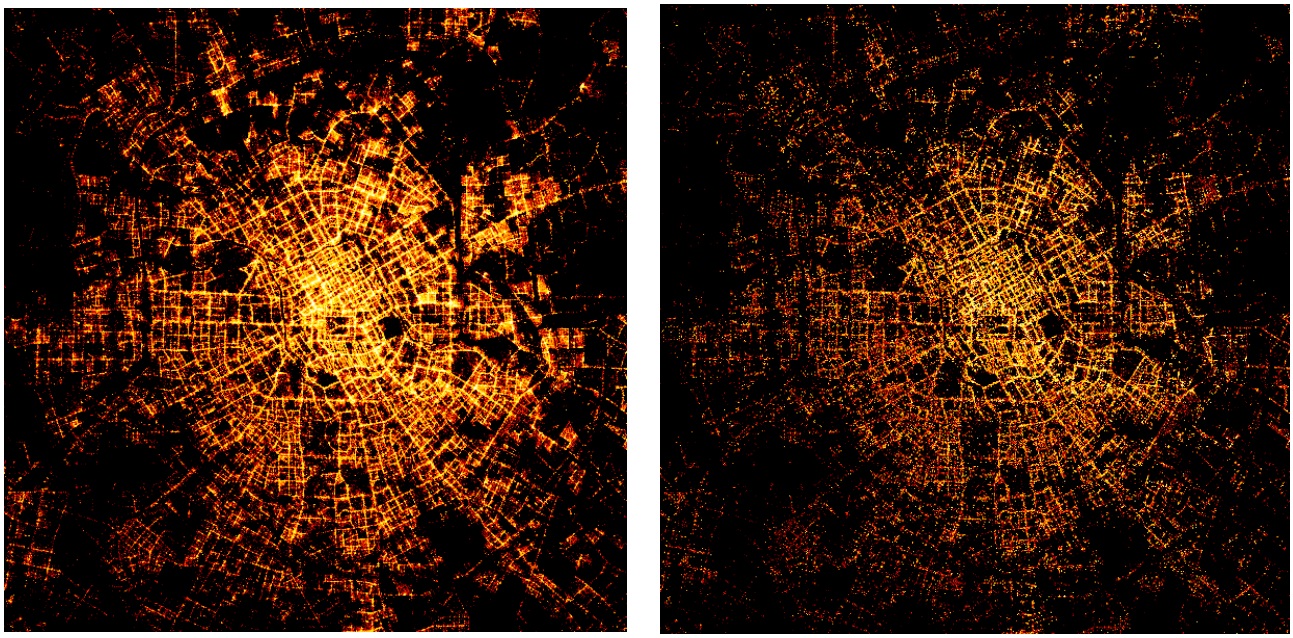

Figure 3 : Pick-ups (left) and drop-offs (right) distribution of DiDi rides in Chengdu city. Color intensity represents the scale of total rides around the location.

4.4.2 Spatiotemporal analysis

We continue by using different dimensions in the data. In this example, we would like to see if certain areas (or roads) are more likely to have pickups at certain hours, a pattern that may origin from natural different urban characteristics of a large city areas, such as night-life areas, residential neighborhoods, business centers or industrial areas. The results, as shown in figure 4, seems to correlate with previous temporal analysis done by Li, Wenxiang, et al [4], as the colors from cyan (12pm), blue (4pm) to purple (8pm) indicates the peak hours off pick-ups. An additional layer unravels when noticing the different colors in the road structure, as more blue shades cover more major roads, what may suggest a later peak in departures from major roads. As for drop-offs, a clearer view is presented, when purple color dominates most of the city structure, suggesting the even later peak (8pm) in most arrivals around the city outer rings. While cyan and green colors (8am-12pm) show up in specific spots in the inner city center, which correlates with the departures morning peak hours.

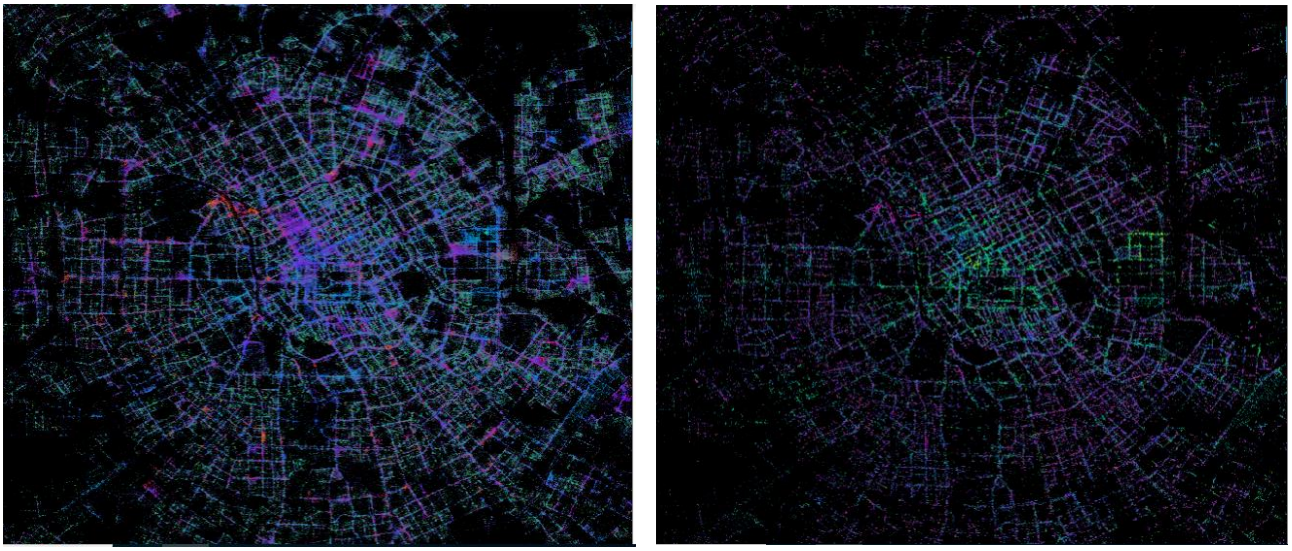

Figure 4: Spatiotemporal representation of DiDi rides pick-ups (left) and drop-offs (right). Since hours and colors are both cyclic, in these visualizations the order of colors is roughly red (midnight), yellow (4am), green (8am), cyan (noon), blue (4pm), purple (8pm), and back to red.

4.4.3 Pick-ups Vs drop-offs trend analysis

To gain additional clear insights regarding urban characteristics of the rides, we plot the pick-ups simultaneously with drop-offs. This results in the following: Roads with more pickups than drop-offs will appear in red scale, while others with more pick-ups will appear in a blue scale. By reviewing figure 5 it is clear that pickups are more common on major roads, as and drop-offs are more common on minor streets, possibly residential. This trend seems even clearer if we examine the most inner ring around the city center.

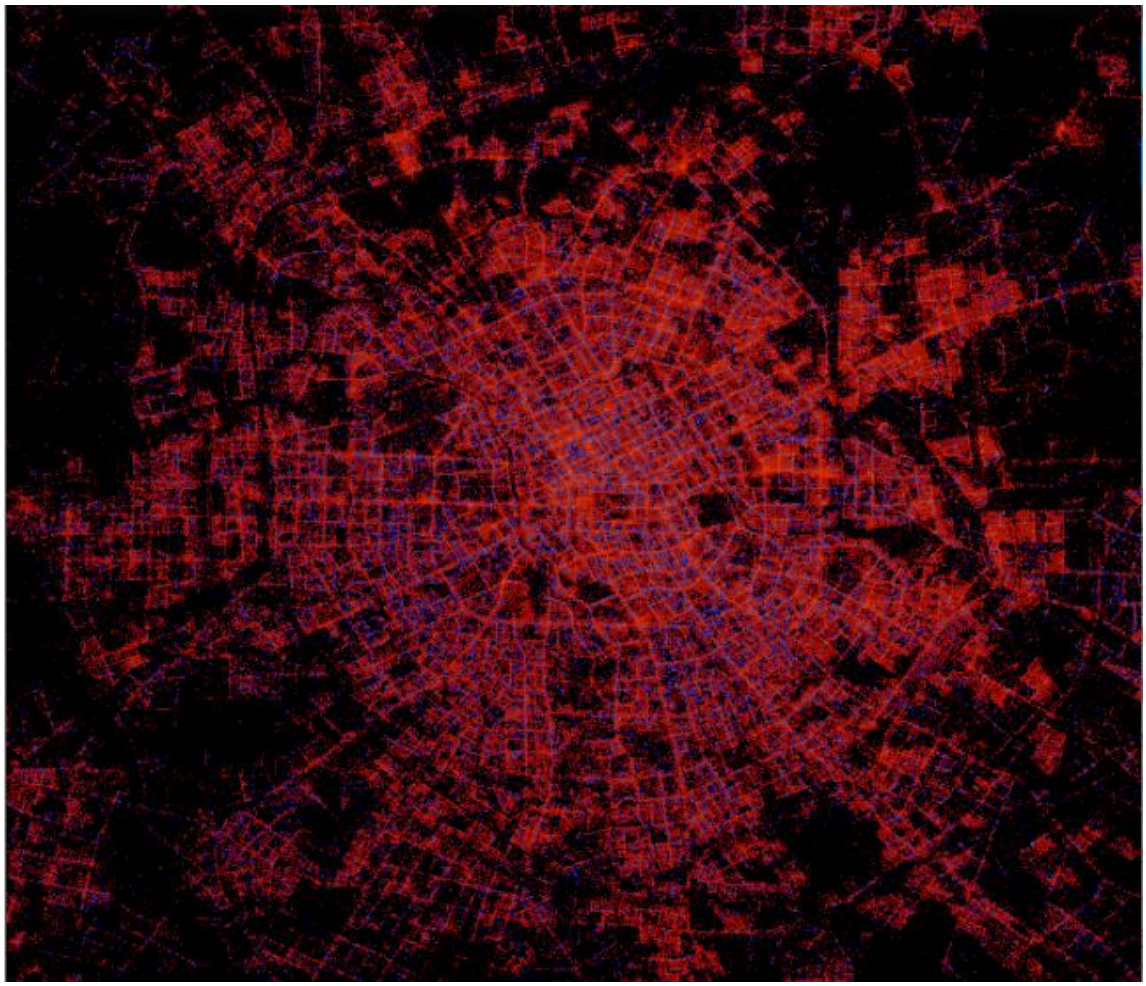

Figure 5 : Pick-ups Vs drop-offs merged ditributions in Chengdu City. Roads with more pick-ups than drop-offs will be colored in red scale, while roads with more drop-offs will be colored in blue scale.

4.5 Hypotheses definition

There is no doubt that many questions arise from these plots, that many of them will require to dig deeper into temporal dimensions (weekdays and hours) and spatial or urban characteristics (points of interest and land-use information). Since our work is limited we chose to focus on the road hierarchy and its behavior that arise from the DiDi data, and other complementary datasets. We define in a more accurately three research questions that we will try to answer following the visual pick-ups/drop offs behavior within city center:

1. Is there a statistical significant change between road length distribution and pick-ups/drop-offs distributions, in relation to road hierarchy distributions?
2. Is there a statistical significant change between pick-ups and drop-offs, in relation to road hierarchy distributions?
3. Is there a statistical significant change between single and shared rides, in relation to road hierarchy distributions, regarding shared rides only? In pick-ups and/or drop-offs.

4.6 Road to rides pick-ups/drop-offs matching

Since the available dataset does not provide the road information we need we first develop an algorithm (figure 6) to detect the nearest road to each pick-up/drop-off point, and then extract it's meta-data we need (e.g road level, road length, distance from point). For optimization reasons, around each point we'll create a small radius to select the closest roads to it as candidate roads, and extract the closest of them to the point. In case no roads found in the selected radius, the radios will double its size, until reaching a maximum of 1k, since a minority of the OSM graph edges are defined by relatively distant nodes from each other. Since time complexity per location is still relatively high we choose to apply the algorithm on 1 week of data only (1.14M records, 18.7% of total), using the *dask* python library for CPU parallelization [16].

get_closest_road (latitude, longitude, osm_graph):

1. *Create a graph from driving roads network within 10KM radius from city center*
2. *For each p in coordinate points [pickups U drop-offs]:*
    a. *r = 50m // initializing minimal radios around point*
    b. *While r <= 1k:*
        i. *p_bbox = Truncate bounding box around point p in radius r*
        ii. *roads = all roads found within cp_bbox*
        iii. *if roads not empty:*
            1. *for road in roads:*
                a. *distances[road] = point_to_line(p, road) [23]*
            2. *closest_road = argmin(distances) // find road with minimum distance*
            3. *return closet road details (highway_type, distance from p)*
        iv. *r = r*2 // expand search parameter around point*

Figure 6: Pseudo-code of road to point matching algorithm

In order to verify the integrity of our results we extracted the distribution of the distances between our inputs (coordinate points) and outputs (closest roads). The results show the median distance stands on 6m, with a maximum of 210m, which are reasonable based on GPS accuracy. Full distribution is detailed in table 5.

Table 5: Distance distributions of pick-up point to nearest road, as found by the 'road to point' matching algorithm

Percentile	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**max**
Distance from nearest road (m)	0.9	1.9	3.1	4.6	6.3	9.4	15	26.3	51.5	210

4.7 City road hierarchy outline

We continue to explore the characteristics of the city road hierarchy and by utilizing the full diversity of Chengdu city. In figure 7 we plot the rides pick-ups and drop-offs, while coloring the different road levels discovered by the road matching methodology we implemented. The plot gives an interesting and colorful overview of the urban characteristics of the city, revealing the surprising density of residential roads (white) and the relations between primary (red), secondary (purple) and tertiary (orange) roads, covering the rings of city center and the major roads. This suggest two conclusions: 1) Most road levels are distributed all around the city center, so any future ride policy that may be suggested is not likely to ignore any part of the city. 2) Current DiDi driving policy is not ignoring any road level, as we can see all 8 colors, representing even the most major road (motorway; yellow) all the way down to the lowest level (living street; blue). Saying that, it seems that the coverage of these road levels, at least based on pick-ups and drop-offs, is very low. In the next section we will explore how low or high is the coverage of different levels, in a more normalized method.

Figure 7: Chengdu city road hirarchy outline, as emerging from DiDi rides pick-ups (left) and drop-offs (right).

***Color legend:** Motorway – yellow; Trunk – Pink; Primary – Red; Secondary – Purple; Tertiary – Orange; Unclassified – Grey; Residential – While; Living Street - Blue*

5 Results

5.1 Statistical analysis of rides behavior

After summarizing our results (see table 6) we'll try to test our hypothesis in statistical methods, which will back-up our visual insights. For each hypothesis, we'll start by analyzing the gaps between distributions, and later test our hypothesis using Pearson Chi-squared test for categorical data.

Table 6: Chengdu road level distribution by city length (left), DiDi rides pick-ups (middle) and drop-offs (right). Depatures and arrivals distributions are also split to single and shared rides, as extracted in sec 6.1. Each column represent a different distribution. Road levels are ordered descending from most major to most minor.

	Road Length		Pickups			Drop-offs		
Road Level	Total Length (KM)	Share of Total Length	Level Share - All rides	Level share - Single Rides	Level share - Shared Rides	Level Share - All rides	Level share - Single Rides	Level share - Shared Rides
motorway	139.7	2.4%	0.17%	0.17%	0.21%	0.44%	0.45%	0.38%
trunk	397.3	6.9%	2.20%	2.20%	2.14%	2.43%	2.45%	2.24%
primary	629.3	10.9%	24.22%	24.10%	25.93%	24.28%	24.37%	23.03%
secondary	761.7	13.2%	17.18%	17.15%	17.59%	15.71%	15.72%	15.52%
tertiary	1466.6	25.4%	27.20%	27.27%	26.20%	24.69%	24.61%	25.78%
unclassified	613.3	10.6%	8.57%	8.56%	8.72%	9.41%	9.45%	8.93%
residential	1673.3	28.9%	19.59%	19.67%	18.38%	21.86%	21.78%	23.00%
living street	103.6	1.8%	0.88%	0.88%	0.82%	1.16%	1.17%	1.12%

- Road length vs all pick-ups: When comparing the road length distribution to entire rides pick-ups distribution we notice the following:
    - City length distribution: Residential roads account for most of road length (28.9%) followed by tertiary (25.4%) and Secondary (13.2%)
    - Major roads: Primary/Secondary/Tertiary share is larger than their relative city length share (68.5% vs 49.4%)
    - Minor roads: Unclassified/Residential/Living street share is smaller than their relative city length share (28.9% vs 41.3%)
    - Chi-Squared test: We reject the Null hypothesis with p-value approaching 0, meaning the distributions are independent and different by any confidence level.
- Pick-ups Vs drop-offs: when comparing the pick-ups distribution vs drop-offs distribution we also learn that:
    - Pick-ups are more popular in tertiary (+2.6% than drop-offs) and secondary (+1.4%) roads, which are major, and less popular in Residential (-2.3%) and unclassified (-0.9%) roads, which are minor.
    - Chi-Squared test: We reject the Null hypothesis with p-value approaching 0, meaning the distributions are independent and different by any confidence level.
- Shared Vs single rides: when comparing shared rides vs single the trend is somewhat mixed:

- Shared pick-ups are more popular in primary (+1.9% than single pick-ups) and secondary (+0.5%) roads, which are major, but are less popular in tertiary (-1%), which is also major and in residential (-1.3%) which is minor.
- Chi-Squared test (pick-ups): We reject the Null hypothesis with p-value approaching 0, meaning the distributions are independent and different by any confidence level.
- Shared drop-offs are more popular in tertiary (+1.1% than single drop-offs) and residential (+1.2%) roads, but are less popular in primary (-1.3%), which and in unclassified (-1.3%) which is minor.
- Chi-Squared test (drop-offs): We reject the Null hypothesis with p-value approaching 0, meaning the distributions are independent and different by any confidence level.

5.2 Temporal analysis of rides road hierarchy behavior

We continue to explore trends in a temporal dimension. To understand the popularity of departure and arriving in different road levels we split the data to weekdays (Mon-Fri) Vs weekends (Sat-Sun), plotting an hourly heat-maps for pick-ups and for drop-offs (figure 8). Taking into consideration the different scale ranges between each level, we normalize each level to its maximum value, so each row in a given heat-map stands on its own.

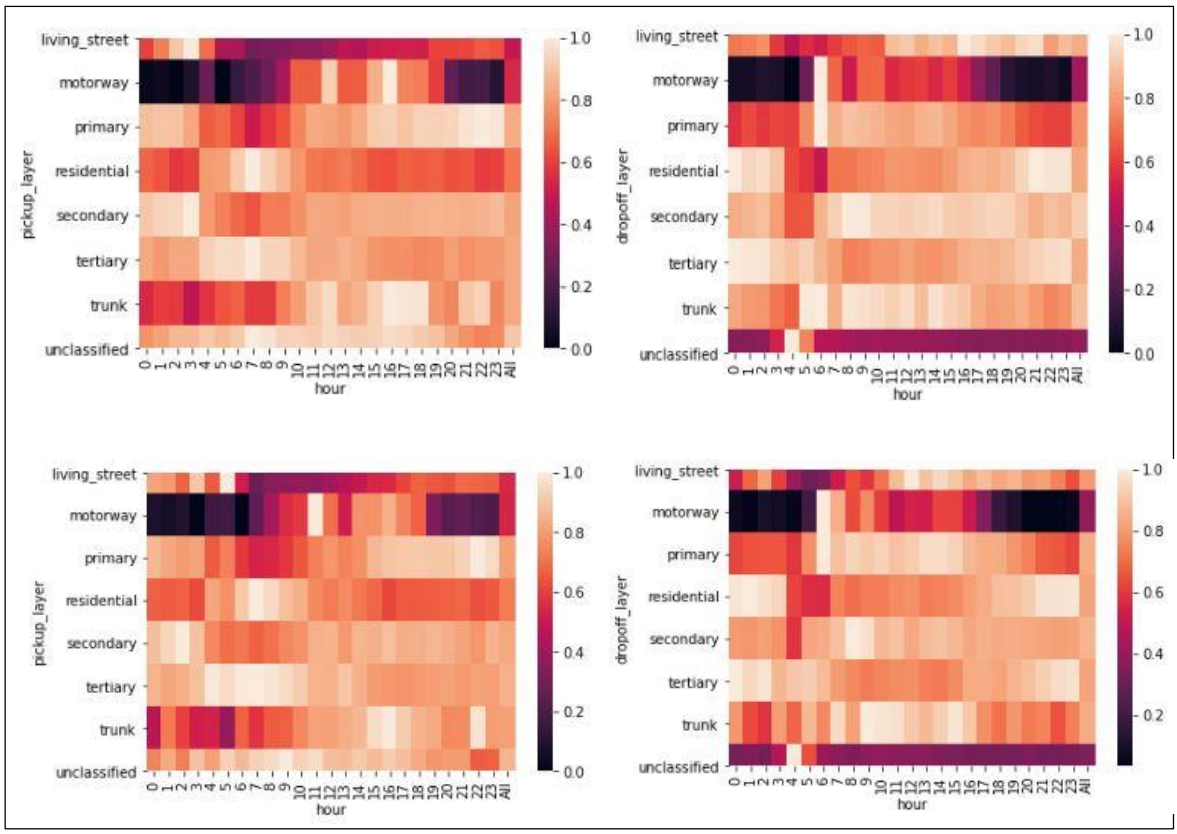

Figure 8: Hourly distributions of Chengdu road level across rides in weekdays: pick-ups (upper-left), and drop-offs (upper right), and in weekends: pick-ups (lower-left) and drop-offs (lower-right). Each road level is normalized by its maximum value.

From this plot we summarize a few insights regarding the more significant levels we focused so far:

- Overall behavior off week-days vs weekends doesn't show any dramatic change, emphasizing the high mobility of the population thought the entire week and the necessity of mobility solutions.
- Pick-ups from primary roads peaks in late evening, while drop-offs peaks at early morning. During weekends drop-offs early peak lasts until afternoon. (probably related to business areas)
- Pick-ups from residential roads peaks in morning, while drop-offs peaks at night. During weekends drop-offs late peak lasts up to 2AM (probably related to night-life habits).
- Temporal behavior of rides from/to tertiary roads tend to be more similar to residential roads, which is interesting, as tertiary relates to more major roads. According to this, it would be interesting for future research to explore the land-use and points of interest around primary vs tertiary roads, which behave in contrary to one another.

5.3 . Visual analysis of rides roads origins to destinations flow

As a sub-task for our goal to understand the rides behavior in relation to the city road hierarchy, we further now try to get insights regarding the flow of rides across the diverse road hierarchy in Chengdu. By understanding not only the behavior of pick-ups or drop-offs roads separately, but the distribution of both for each ride, we can point to the more and less "common" journeys across the city, in relation to road levels. For each ride we'll use both road labels previously attached to it (pick-up and drop-off), and analyze the distribution of origin level to destination level using the Sankey diagram [21].

By analyzing the results from diagram (figure 9) and the data table created it (appendix 1) we can gain a few insights:

- The distribution of destinations road levels within each origin road level group is relatively similar to the distribution of destinations across all rides, maintaining the most popular level as tertiary, followed by primary, residential, secondary, unclassified, living street and motorway as the least popular destination.
- In contrary, the distribution of pick-ups road levels within each destination road level group is relatively similar to the distribution of pick-ups across all rides.
- Intra higher level Vs intra lower level: 47.8% of rides travel within (from and to) major roads (tertiary and higher) only while only 9.3% or rides travel within minor roads (unclassified and lower). The rest, 42.9% are crossing higher and lower levels.

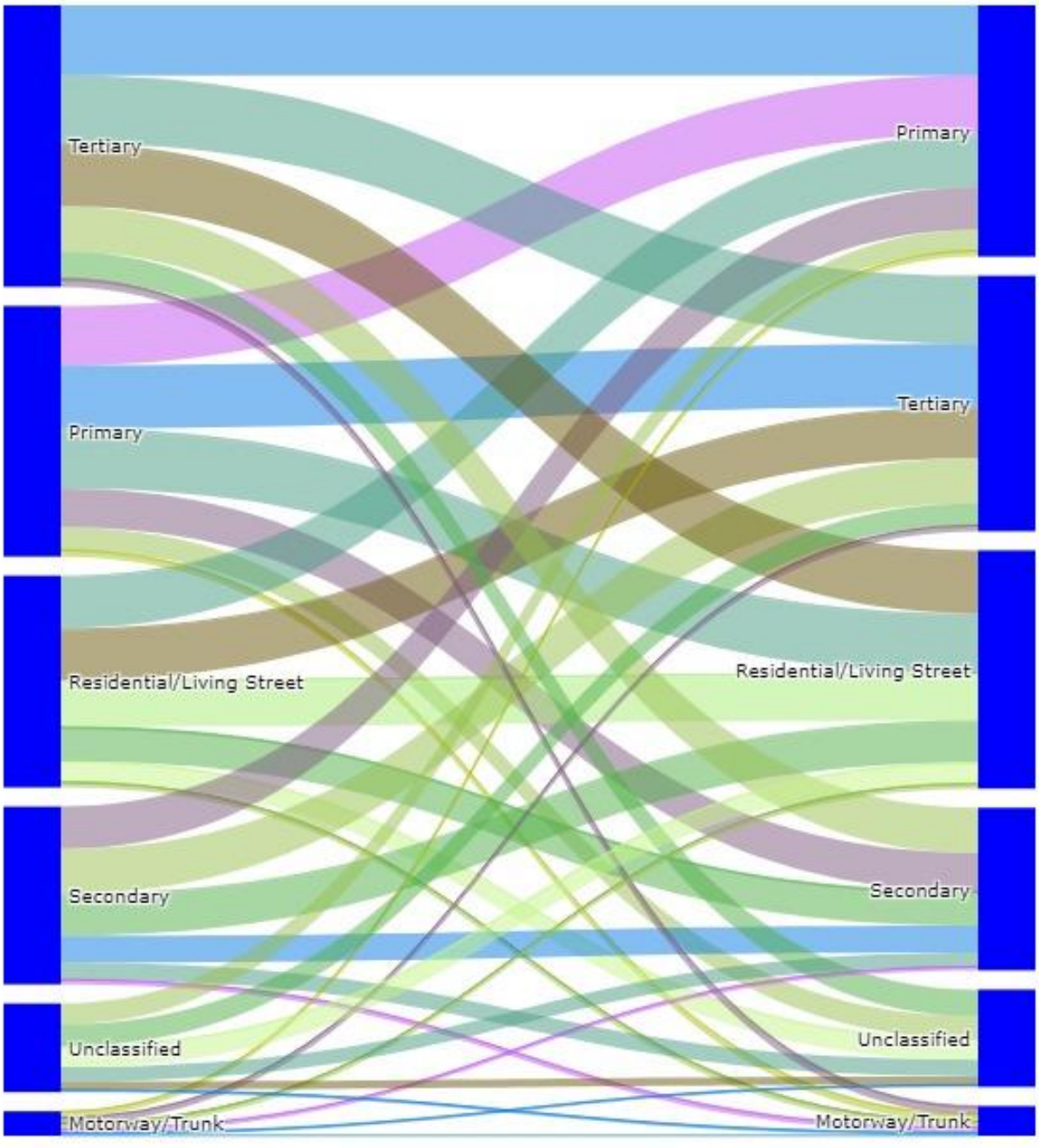

Figure 9: Flow distribution of road levels from DiDi rides pick-ups (left) to drop-offs (right).
For a better visual experiense, small share road levels, such as motorway, trunk and living street, were grouped to their closest hierarchy group

6 Discussion and future work

Chengdu, a capital city in southwest China with an urban population of 11.2 million, relies heavily on public transportation including mobility on-demand services, with DiDi being the most popular service for ride-sourcing and ride-sharing, with more than 8.5 million users. As an emerging sub-category of MoD services, the significant societal and environmental benefits of ride-sharing has been deeply researched and highly discussed in the last few years. The potential in megacities such as New York is enormous, as it has been found that percentage of shared trips in the city can potentially be increased from 7.85% to 90.69%, and the percentage of time savings can reach 25.75% from 2.38%.

A relatively less studied urban characteristic, the city road hierarchy, reveals itself from the data driven case study in Chengdu for the first time. First, it reveals itself theoretically from OSM data, and later in practice from DiDi rides pick-ups and drop-offs. It is found to be highly diverse, covering vast majority of known driving network of city center and comprised of eight different levels. Primary, Tertiary and residential roads found to be covering the majority of departures and arrivals across the city, as well as the city road network measured length.

Further analyzing this urban layer using available taxi data from DiDi and complementary open source data, visual and statistical methods presents different distributions of rides departures vs arrivals, with relation to their road levels in the city hierarchy. When taking into consideration the large scale of rides in our sample, the change in distributions shows significant tendency toward major roads in ride pick-ups, and towards minor road in pick-ups. A significant distributions change between road level popularity discovers as well when comparing single vs shared rides, which take up to 6-7% from the studied dataset. Saying that, the trends in this sub-category are less clear in relation to major and minor roads, both across pick-ups and drop-offs. In addition, the temporal characteristics of the different city road levels reveals a wide set of different behaviors, which among them are different peak days, hours and durations across weekdays and weekends.

When considering all the findings above it is clear that the characteristics of the city road hierarchy, such as its outline, its multi-level diversity and its spatiotemporal popularity trends - we believe that this feature can and should be a part of any smart ride-service policy design. Emerging smart ride-sharing services, such as Via, MOIA and others who applies corner-to-corner pick-up and drop-off policies, can especially benefit from leveraging knowledge embedded within city road hierarchy, on top of all other online city and service parameters. Considering that avoiding specific types of roads, whether minor- residential or major outside of urban areas, are part of these services policies main features and distinction - the knowledge and characteristics of these roads should be maximized in order to better achieve its policy goals. Other ride-sharing companies, such as DiDi and others, can benefit from leveraging this knowledge when considering introducing corner-to corner services or similar.

Our method, results and conclusions are presented above, but is limited by many aspects and research gaps, that could be filled by future work. The main limitations of this work and its conclusions are presented here.

Considering a city with urban population of 11.2 million and 5 million cars, any results regarding trends in the behavior of specific road level(s) we should be aware that our results were not merged with additional urban layers, such as POI data of Chengdu, population density, land-use data or public transit stops. These layers were explored by other DiDi related research, but not combined so far with city road hierarchy layer, as far as we know. Mixing those layers could add additional information regarding findings. In addition, it's noticeable that some aspects of the DiDi dataset were not leveraged in this work. These range from original aspects such as trajectory data, (which found to be incomplete and needed reconstruction) and others such as ride duration, air distance, actual distance, evaluated route and evaluated speed – all can be extracted using orders and trajectory data manipulation. Analyzing ride actual route in relation to road hierarchy, for example, could shed some light on more than the departure and arrival roads.

From a data integrity view, we understand the limitation of open source projects. Our main complementary dataset merged with the DiDi data to create the city road hierarchy layer is open street maps, which relies on community contribution. As studied by Zielstra et al. [20], while this data import built the foundation for the active OSM community, the poor quality of the imported road data and additional problems with their conversion to the OSM tagging scheme limited OSM data usability and reliability. Based on this, it would be interesting to reproduce the results using more official dataset, even though the OSM data of Chengdu specifically found as relatively versatile, which generally implies towards completeness and accuracy.

Finally, as we tried to achieve our second goal in this work, The DiDi dataset provides GPS accuracy of rides pick-ups and drop-offs, hence the users walking/driving distance from their original departure/arrival addresses are not available to us. Additional data holding this information, at least for shared rides, will shed light on the potential of corner-to-corner policy applications. Additional features of the vehicles such as passenger capacity and activity hours could help for future research as well.

References

[1] Gao, Qingke, et al. "Identification of urban regions' functions in Chengdu, China, based on vehicle trajectory data." PloS one 14.4 (2019): e0215656.

[2] Kong, Hui, Xiaohu Zhang, and Jinhua Zhao. "How does ridesourcing substitute for public transit? A geospatial perspective in Chengdu, China." Journal of Transport Geography 86 (2020): 102769.

[3] Tu, Meiting, et al. "Improving ridesplitting services using optimization procedures on a shareability network: A case study of Chengdu." *Technological Forecasting and Social Change* 149 (2019): 119733.

[4] Li, Wenxiang, et al. "Characterization of ridesplitting based on observed data: A case study of Chengdu, China." *Transportation Research Part C: Emerging Technologies* 100 (2019): 330-353.

[5] https://towardsdatascience.com/if-taxi-trips-were-fireflies-1-3-billion-nyc-taxi-trips-plotted-b34e89f96cfa

[6] Alonso-Mora, Javier, et al. "On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment." *Proceedings of the National Academy of Sciences* 114.3 (2017): 462-467 .

[7] Schrank D, Eisele B, Lomax T (2012) Texas Transportation Institute 2012 Urban Mobility Report (Texas Transportation Institute, A&M University, College Station, TX).

[8] Pant P, Harrison RM (2013) Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: a review. Atmos Environ .97–77:78

[9] Carrion C, Levinson D (2012) Value of travel time reliability: a review of current evidence. Transp Res Part A Policy Pract 46(4):720–741.

[10] Hennessy DA, Wiesenthal DL (1999) Traffic congestion, driver stress, and driver aggression. Aggress Behav 25(6):409–423.

[11] Yu, Biying, et al. "Environmental benefits from ridesharing: A case of Beijing." Applied energy 191 (2017): 141-152.

[12] Zhang, Yingjia, et al. "Density and diversity of OpenStreetMap road networks in China." *Journal of Urban Management* 4.2 (2015): 135-146.

[13] https://en.wikipedia.org/wiki/Restrictions_on_geographic_data_in_China#GCJ-02

[14] https://github.com/googollee/eviltransform

[15] https://wiki.openstreetmap.org/wiki/Key:highway

[16] https://docs.dask.org/en/latest/

[17] https://osmnx.readthedocs.io/en/stable/index.html

[18] https://datashader.org/

[19] Jilani, Musfira, Padraig Corcoran, and Michela Bertolotto. "Automated highway tag assessment of OpenStreetMap road networks." *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2014.

[20] Zielstra, Dennis, Hartwig H. Hochmair, and Pascal Neis. "Assessing the Effect of Data Imports on the Completeness of OpenStreetMap–AU Nited S Tates Case Study." *Transactions in GIS* 17.3 (2013): 315-334.

[21] https://en.wikipedia.org/wiki/Sankey_diagram

[22] https://outreach.didichuxing.com/research/opendata/en/#

[23] https://mathworld.wolfram.com/Point-LineDistance2-Dimensional.html

[24] https://www.businessofapps.com/data/uber-statistics/

[25] https://www.businessofapps.com/data/lyft-statistics/

[26] https://en.wikipedia.org/wiki/Via_(company)

[27]https://modo.volkswagengroup.it/en/q-life/moia-the-shared-mobility-of-the-future-is-already-here

[28] Gilibert, Mireia, et al. "Mapping of service deployment use cases and user requirements for an on-demand shared ride-hailing service: MOIA test service case study." *Case Studies on Transport Policy* 7.3 (2019): 598-606.

Appendix 1

Pick-ups road to drop-offs road flow distribution, grouped by pick-up road level

pickup_level	dropoff_level	all_rides	single_rides	shared_rides
living_street	living_street	1.2%	1.2%	0.3%
living_street	motorway	0.4%	0.4%	0.3%
living_street	primary	23.2%	23.4%	20.9%
living_street	residential	23.8%	23.7%	26.2%
living_street	secondary	15.6%	15.7%	14.8%
living_street	tertiary	24.1%	23.9%	26.5%
living_street	trunk	2.2%	2.2%	2.1%
living_street	unclassified	9.5%	9.5%	8.8%
motorway	living_street	1.0%	0.9%	1.2%
motorway	motorway	0.3%	0.3%	0.0%
motorway	primary	23.2%	23.9%	14.8%
motorway	residential	21.8%	21.8%	21.6%
motorway	secondary	15.5%	15.2%	18.5%
motorway	tertiary	25.9%	25.4%	32.1%
motorway	trunk	3.5%	3.6%	2.5%
motorway	unclassified	8.9%	8.8%	9.3%
primary	living_street	1.1%	1.2%	1.1%
primary	motorway	0.4%	0.4%	0.3%
primary	primary	23.7%	23.9%	21.6%
primary	residential	22.9%	22.8%	24.4%
primary	secondary	15.4%	15.4%	15.1%
primary	tertiary	25.1%	25.0%	27.0%
primary	trunk	2.4%	2.4%	2.3%
primary	unclassified	8.9%	9.0%	8.3%
residential	living_street	1.2%	1.2%	1.2%
residential	motorway	0.5%	0.5%	0.4%
residential	primary	24.9%	24.9%	24.1%
residential	residential	20.8%	20.7%	21.7%
residential	secondary	16.1%	16.1%	15.8%
residential	tertiary	24.4%	24.3%	24.9%
residential	trunk	2.3%	2.3%	2.2%
residential	unclassified	9.9%	9.9%	9.8%

pickup_level	dropoff_level	all_rides	single_rides	shared_rides
secondary	living_street	1.2%	1.2%	1.1%
secondary	motorway	0.4%	0.4%	0.4%
secondary	primary	23.8%	23.9%	22.9%
secondary	residential	22.4%	22.3%	23.4%
secondary	secondary	15.1%	15.1%	15.0%
secondary	tertiary	25.2%	25.1%	26.3%
secondary	trunk	2.4%	2.4%	2.3%
secondary	unclassified	9.5%	9.5%	8.7%
tertiary	living_street	1.1%	1.1%	1.1%
tertiary	motorway	0.5%	0.5%	0.5%
tertiary	primary	24.8%	24.9%	24.1%
tertiary	residential	21.3%	21.2%	22.2%
tertiary	secondary	16.0%	16.0%	15.8%
tertiary	tertiary	24.2%	24.1%	24.7%
tertiary	trunk	2.5%	2.5%	2.3%
tertiary	unclassified	9.6%	9.6%	9.3%
trunk	living_street	1.0%	0.9%	1.8%
trunk	motorway	0.6%	0.6%	0.4%
trunk	primary	24.6%	24.6%	23.9%
trunk	residential	21.3%	21.3%	20.8%
trunk	secondary	15.4%	15.4%	16.1%
trunk	tertiary	25.5%	25.6%	24.7%
trunk	trunk	2.8%	2.8%	1.9%
trunk	unclassified	8.9%	8.8%	10.5%
unclassified	living_street	1.2%	1.2%	1.3%
unclassified	motorway	0.4%	0.5%	0.3%
unclassified	primary	23.7%	23.8%	22.3%
unclassified	residential	22.1%	22.0%	23.5%
unclassified	secondary	16.1%	16.1%	16.1%
unclassified	tertiary	24.7%	24.6%	26.0%
unclassified	trunk	2.5%	2.5%	2.2%
unclassified	unclassified	9.2%	9.3%	8.3%

Appendix 2

Pick-ups road to drop-offs road flow distribution, grouped by drop-off road level

pickup_level	dropoff_level	all_rides	single_rides	shared_rides	pickup_level	dropoff_level	all_rides	single_rides	shared_rides
living_street	living_street	0.9%	0.9%	0.2%	secondary	living_street	0.9%	0.9%	0.8%
living_street	motorway	0.1%	0.1%	0.2%	secondary	motorway	0.2%	0.2%	0.3%
living_street	primary	23.8%	23.8%	24.5%	secondary	primary	23.7%	23.6%	25.2%
living_street	residential	20.6%	20.7%	19.0%	secondary	residential	20.1%	20.2%	18.7%
living_street	secondary	17.9%	17.9%	17.1%	secondary	secondary	16.5%	16.4%	17.0%
living_street	tertiary	26.1%	26.2%	25.3%	secondary	tertiary	27.7%	27.8%	26.8%
living_street	trunk	1.9%	1.8%	3.4%	secondary	trunk	2.2%	2.1%	2.2%
living_street	unclassified	8.7%	8.6%	10.2%	secondary	unclassified	8.8%	8.8%	9.1%
motorway	living_street	0.8%	0.8%	0.7%	tertiary	living_street	0.9%	0.9%	0.8%
motorway	motorway	0.1%	0.1%	0.0%	tertiary	motorway	0.2%	0.2%	0.3%
motorway	primary	21.9%	21.9%	22.1%	tertiary	primary	24.6%	24.4%	27.2%
motorway	residential	20.3%	20.3%	20.1%	tertiary	residential	19.3%	19.4%	17.8%
motorway	secondary	15.2%	15.1%	16.3%	tertiary	secondary	17.5%	17.5%	17.9%
motorway	tertiary	30.1%	30.0%	31.3%	tertiary	tertiary	26.6%	26.7%	25.1%
motorway	trunk	3.0%	3.0%	2.0%	tertiary	trunk	2.3%	2.3%	2.1%
motorway	unclassified	8.7%	8.8%	7.5%	tertiary	unclassified	8.6%	8.6%	8.8%
primary	living_street	0.8%	0.8%	0.7%	trunk	living_street	0.8%	0.8%	0.8%
primary	motorway	0.2%	0.2%	0.1%	trunk	motorway	0.2%	0.2%	0.2%
primary	primary	23.6%	23.6%	24.3%	trunk	primary	24.1%	24.0%	26.5%
primary	residential	20.1%	20.1%	19.2%	trunk	residential	18.6%	18.6%	17.9%
primary	secondary	16.9%	16.8%	17.5%	trunk	secondary	17.2%	17.1%	17.8%
primary	tertiary	27.8%	27.9%	27.4%	trunk	tertiary	28.0%	28.1%	26.5%
primary	trunk	2.2%	2.2%	2.2%	trunk	trunk	2.5%	2.5%	1.8%
primary	unclassified	8.4%	8.4%	8.4%	trunk	unclassified	8.6%	8.6%	8.5%
residential	living_street	1.0%	1.0%	0.9%	unclassified	living_street	0.9%	0.9%	0.8%
residential	motorway	0.2%	0.2%	0.2%	unclassified	motorway	0.2%	0.2%	0.2%
residential	primary	25.4%	25.2%	27.5%	unclassified	primary	23.0%	22.9%	24.0%
residential	residential	18.6%	18.7%	17.3%	unclassified	residential	20.5%	20.5%	20.1%
residential	secondary	17.6%	17.6%	17.9%	unclassified	secondary	17.3%	17.3%	17.1%
residential	tertiary	26.5%	26.6%	25.3%	unclassified	tertiary	27.7%	27.7%	27.2%
residential	trunk	2.1%	2.2%	1.9%	unclassified	trunk	2.1%	2.1%	2.5%
residential	unclassified	8.7%	8.6%	8.9%	unclassified	unclassified	8.4%	8.4%	8.1%