



CAR SHARING IN CHINA

Group member:
XIAO FENG
JINGYAN ZHANG
ALDO MYRTAJ



01

MOTIVATION

Mobile Car-sharing blooms in China

Roland Berger 最新の知見 専門分野 会社概要 採用情報 > Jobs

Think:Act
navigating complexity

CHINESE CAR-SHARING IS BLOSSOMING

TECHINASIA Articles NEW! Jobs Events About Search Like 792K Sign up Log in Write

Didi Kuaidi's rise to power (INFOGRAPHIC)

9 COMMENTS Erik Crouch 9:39 AM at Jan 29, 2016 | 1 min read

101 f t in

It's hard to believe that Didi Kuaidi, a company with more than 250 million users and a value upwards of \$16 billion, is such a new company.

This time last year, in fact, it didn't exist at all. At that point, there were still two competing apps and services, Didi Dache and Kuaidi Dache. The fight was fierce – and in February of 2015, the two decided that if you can't beat 'em, merge with 'em. The merger is helping Didi battle Uber as the US-based startup expands across China.

Need Help? ☰

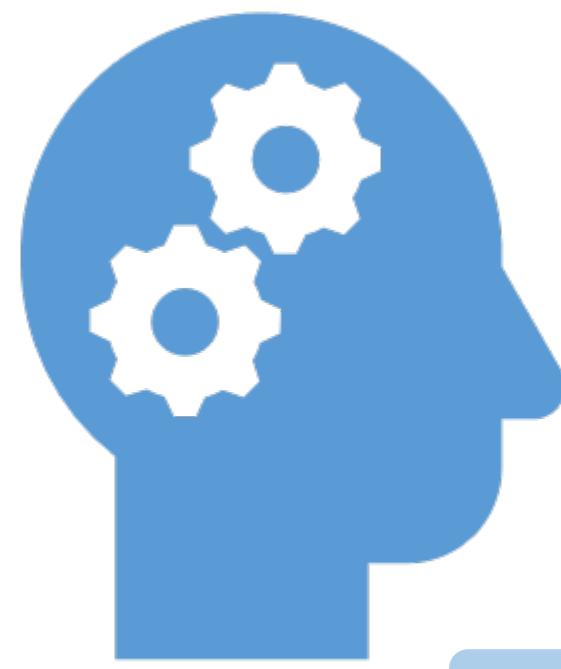
FORTUNE Didi's China Users Can't Stop Complaining About Higher Prices

INTERNATIONAL • CHINA

Ad closed by Google Report this ad AdChoices ▾

Didi's China Users Can't Stop Complaining About Higher Prices

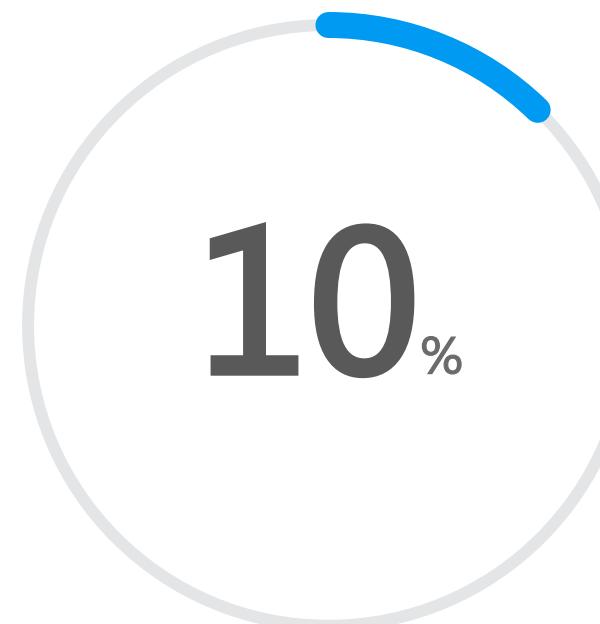
f t in e



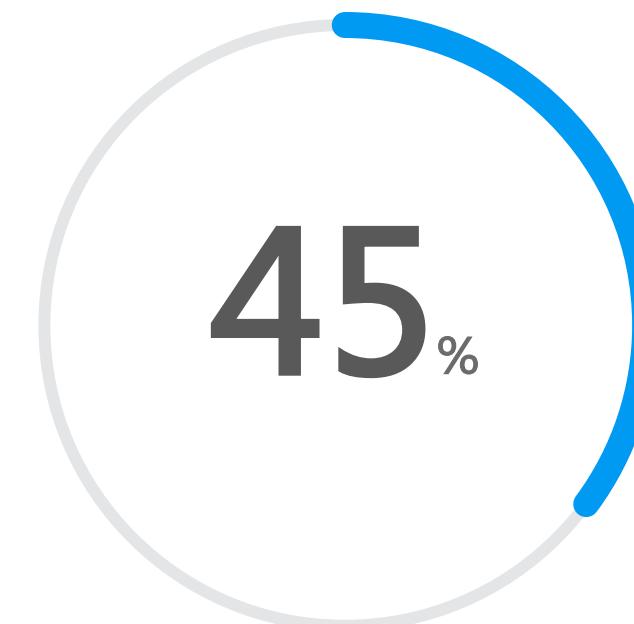
Question Generation

- What is the situation of mobile car sharing in China?
- How people live with mobile car sharing?
- What can we improve in mobile car sharing?

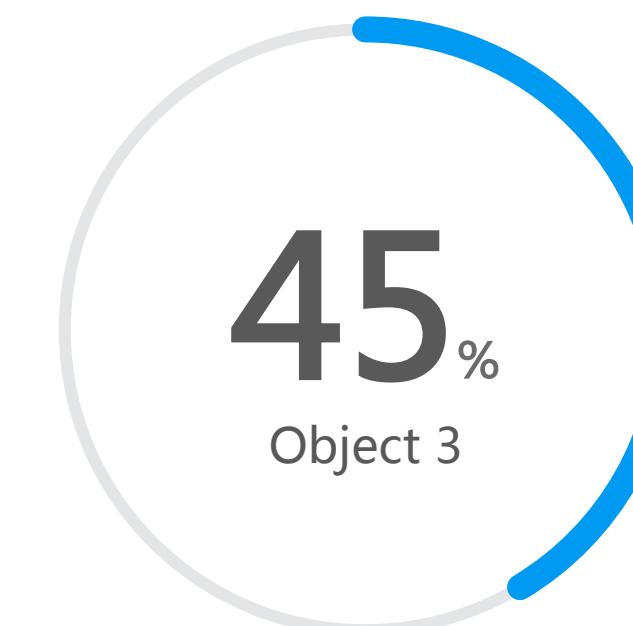
The component of our project



The overview of DIDI in Chengdu



Human life behind DIDI



What we can improve about DIDI
(carsharing inside carsharing)



02

Related work of
CAR SHARING
topic

UBER AND TAXI IN NEW YORK

- Basic analysis:
Time, Duration, Location
- Exploration on location:
neighborhood
- Interesting exploration
on human lifestyle
Nightlife
- Comparison with taxi and
bicycle

Analyzing 1.1 Billion NYC Taxi and Uber with a Vengeance

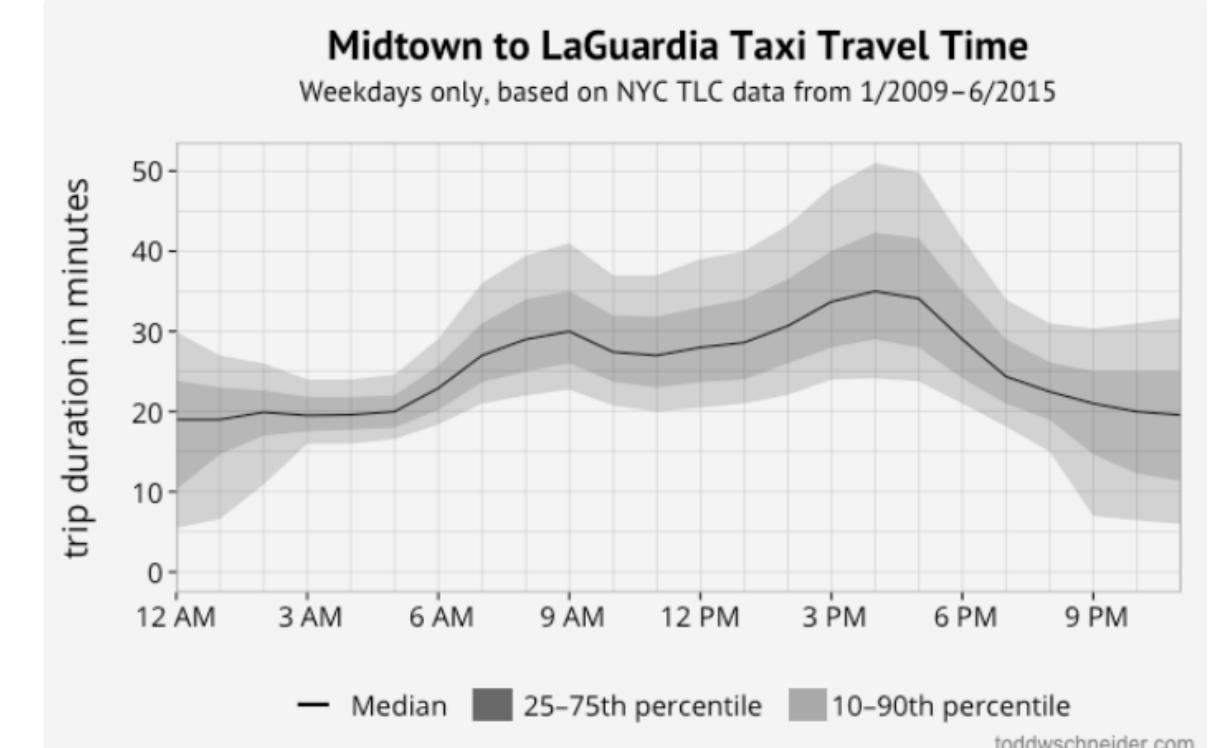
An open-source exploration of the city's neighborhoods, airport traffic, and more, through the lens of publicly available and Uber data

[Tweet](#) [Share](#)

The New York City Taxi & Limousine Commission has released a staggering historical dataset covering over 1.1 billion individual taxi trips in the city from 2009 through June 2015. Taken as a whole, the detailed trip-level data is just a vast list of taxi pickup and drop off coordinates: it's a story of New York. What's bad is the rush hour traffic from Midtown to JFK? Where does the Bridge crowd hang out on Saturday nights? What time do investment bankers leave work? How has Uber changed the landscape for taxis? And could Bruce Willis fit in a limo? Jackson have made it from 72nd and Broadway to Wall Street in less than 10 minutes?

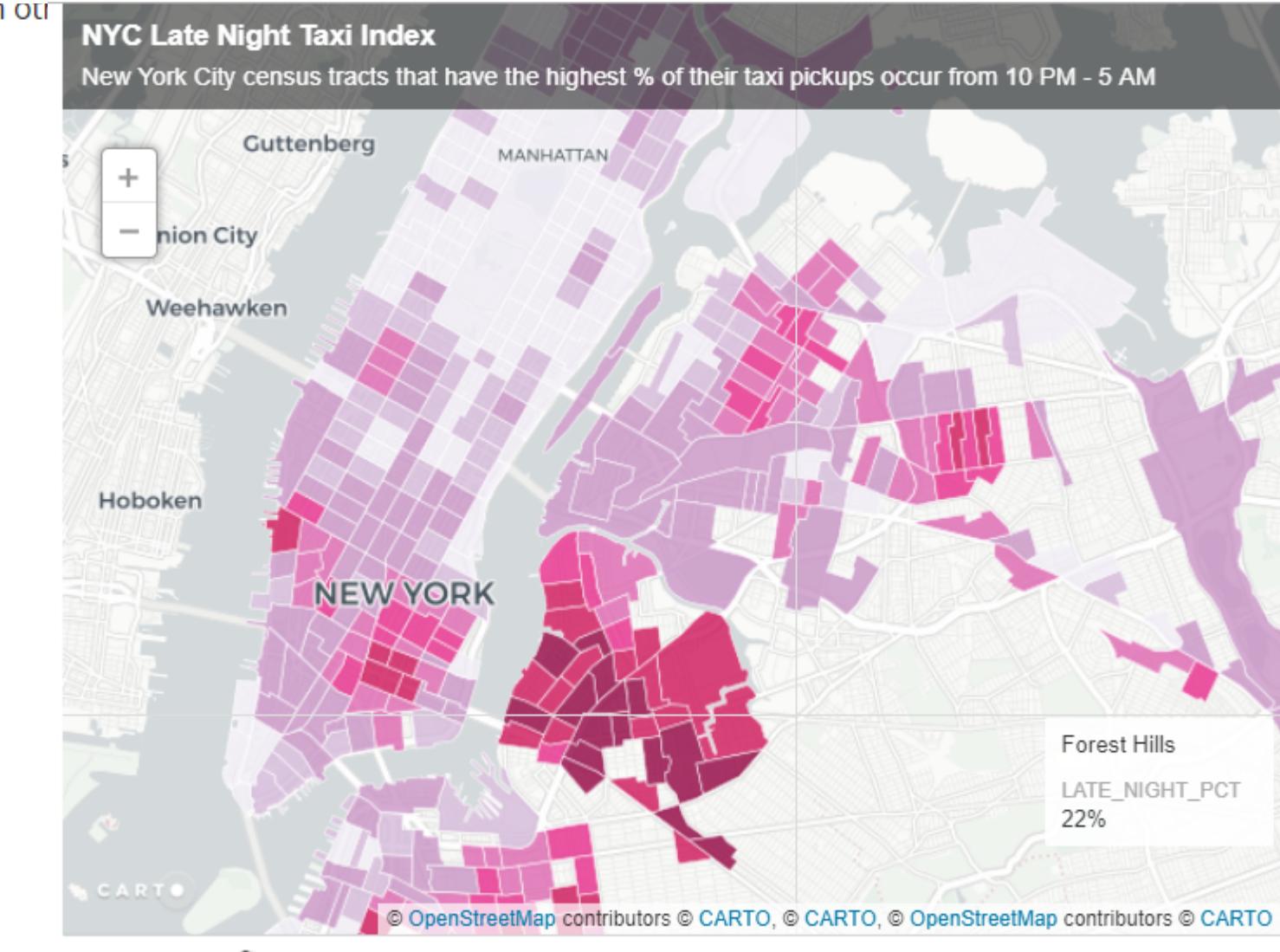
Travel time from Midtown, Manhattan to...

LaGuardia Airport



The stories are similar for traveling to LaGuardia and Newark airports, and from other neighborhoods. You can see the graphs for airport travel times from any neighborhood by selecting it in the dropdown below:

Pick a neighborhood





03

CHINA DIDI DATA AND METHOD

WHAT IS OUR DATASET?

Dataset	Description	Data size	Time period	Data provider
Order data	Order description	800M	2016.11	DIDI CHINA
GPS data	Gps data get from sensor every 3 Minutes	12G	2016.11	DID CHINA

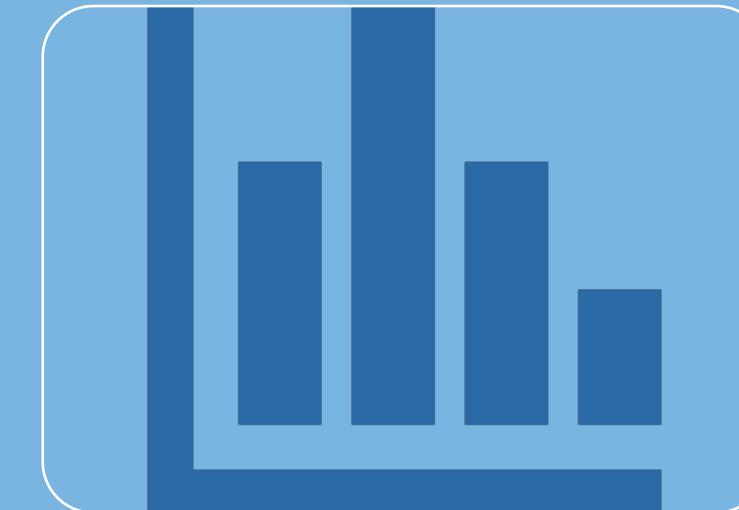
ORDER DATASET

	orderid	stime	etime	slon	slat	elon	elat
1	zfpKH6fweso7znuubokC_gAlmxCwbnwt	1478837401	1478838356	103.970730	30.674890	104.057120	30.671340

GPS DATASET

	carid	orderid	time	lon	lat
1	Bnvzv4pukAo1DmGB0qgB5inJioBv1nri	GnqHA_rjrtt.ttvD.spE.cxClxF_mto	1478838968	104.08284	30.68210

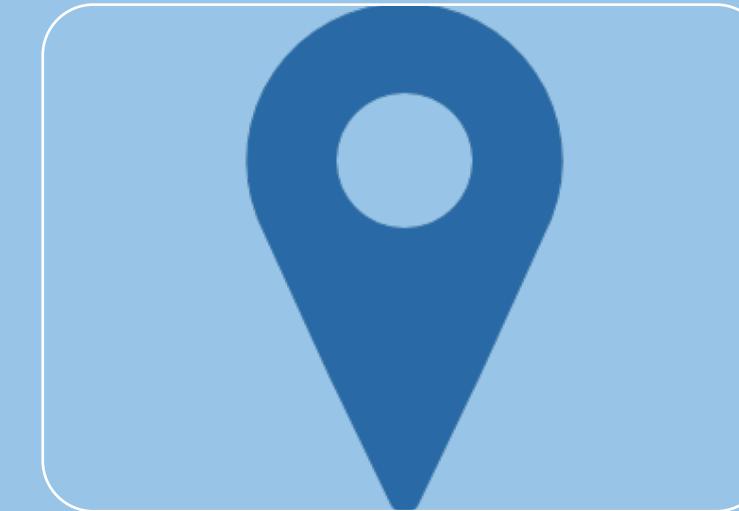
WHAT IS OUR METHOD?



Basic overview

- Bar, line, piechart, density heatmap
- Shapefile

ggplot2, plotly,
rgdal, maptools



How people use DIDI

- STAT2DDENSITY BUILT ON GGMAP LEAFLET
- ANIMATION,
- API FROM REMAP(BAIDU), GOOGLE MAP

Leaflet, ggmap
Magick
REMAP, geosphere



CAR SHARING INSIDE CAR SHARING

- DISTNACE CALCULATION
- HIERARCHICAL CLUSTERING

googleway
geosphere

When and where do people go working, shopping, for entertainment...

Difficulty:

- Offsets among different map systems
- Language Transformation
- Deal with multi order dimensions(time, longitude, latitude, amount...) intuitively
- Explore interesting patterns from large dataset

Our Approach

- Crawl POI data by using map API
 - Offset correction of POI lng/lat
 - Calculate distance between order(start/end) location and POI location
 - Set threshold, transform lng/lat to different type of place
 - Figure out order patterns start and end with particular place (eg. From residential area to office building)
- Spatial Distribution of order start/end on map
 - pick-up/drop-off density on map
 - comparation of high density area with POI markers
- Order distribution with time
 - when and how many people go for working, shopping,...

Main Idea, use an application (algorithm) that suggests users that will travel in the same destination a car/cab they can share at the start.

Difficulty:

- Data Cleaning inconsistent data that did not yield a good result
- Geolocation data takes more processing time in calculations (different transformations)
- Noisy GPS data
- Different from NY data no grouping districts, thus we had to group based on geolocation
- Google API helped us enrich the data though the daily API limit was very fast exceeded (free version no academic licencing)

Our Approach

```
#give as inputs order_data and cuttree for the nr of meters the radius of searching
clustersget_shared_trips <- function(temp, cuttree)
{sp.start <- SpatialPointsDataFrame(temp[, c("slon", "slat")], temp,
proj4string = CRS("+proj=longlat +ellps=WGS84 +datum=WGS84"))
sp.end <- SpatialPointsDataFrame(temp[, c("elon", "elat")], temp,
proj4string = CRS("+proj=longlat +ellps=WGS84 +datum=WGS84"))
mdist <- distm(sp.start)hc <- hclust(as.dist(mdist), method="complete")
sp.start$clust <- cutree(hc, h=cuttree)
mdist <- distm(sp.end)hc <- hclust(as.dist(mdist), method="complete")
sp.end$clust <- cutree(hc, h=cuttree)temp$startclusters <- sp.start$clusttemp$endclusters <-
sp.end$clust
temp$ID <- 1temp <- temp %>% group_by(startclusters, endclusters) %>% filter(ID!
=sum(ID)) %>% select(-ID) %>% ungroup()if(length(as.numeric(row.names(temp))) == 0)
return(0)
temp$group_id <- temp %>% group_indices(startclusters, endclusters)
temp <- temp %>% select(-startclusters, -endclusters)
temp1 <- temptemp1$similartrips <- 1
```

Our Approach

```
temp1 <- temp1 %>% group_by(group_id) %>% summarise(similartrips=sum(similartrips))
temp <- temp[duplicated(temp$group_id), ]
temp <- merge(temp, temp1, by = "group_id")
temp <- temp %>% select(-group_id) return (temp)
}
```

```
get_distance_data <- function(temp){
origin <- list(c(temp[1]$slat,temp[1]$slon))
destination <- list(c(temp[1]$elat,temp[1]$elon))
resp <- google_distance(origins = origin, destinations = destination, mode = "driving",
key = "AIzaSyBii-Ivht8gjIf9fLoMNqVjdd6VTpqcEao",simplify = TRUE)
if(is.null(resp$rows$elements[[1]]$distance$value)) return (0)
temp$distanc <- resp$rows$elements[[1]]$distance$value
temp$duration <- resp$rows$elements[[1]]$duration$value
temp$duration_in_traffic <- resp$rows$elements[[1]]$duration_in_traffic$value return (temp)
}
```

```
# main part calculating everything
p <- 1
temp <- order_data_raw %>% filter(stime >= 1479657610+((p-1)*600) & stime <
1479657610+(p*600))
temp <- get_shared_trips(temp, 500)
shared_trips_1121 <- get_distance_data(temp[1,])
for (p in 2:length(as.numeric(row.names(temp)))) {
  shared_trips_1121 <- rbind( shared_trips_1121, get_distance_data(temp[p,]))
}
# when stuck due to API key please instead of nr 2 input the number that p was when
the loop was stuck (check p in the global enviornment)
for (p in 2:144) {
  temp <- order_data_raw %>% filter(stime >= 1479657610+((p-1)*600) & stime <
  1479657610+(p*600))
  temp <- get_shared_trips(temp, 500)  if(length(as.numeric(row.names(temp))) == 0)
  nextfor (s in 1:length(as.numeric(row.names(temp)))) {
    temp1 <- get_distance_data(temp[s,])
    if(length(as.numeric(row.names(temp1))) == 0) { print("Change the API key")  break }
    shared_trips_1121 <- rbind( shared_trips_1121, temp1)
  }
  print(p)
}
```



04

1st Question: The Situation of DIDI in China

The Overview of DIDI in Chengdu

Car sharing service in china(based on Didi in Chengdu) ≡

- basic Didi situation analysis <
- driver analysis
- POI analysis <

206536times

Average Number of Orders

39595

Average Number of Cars

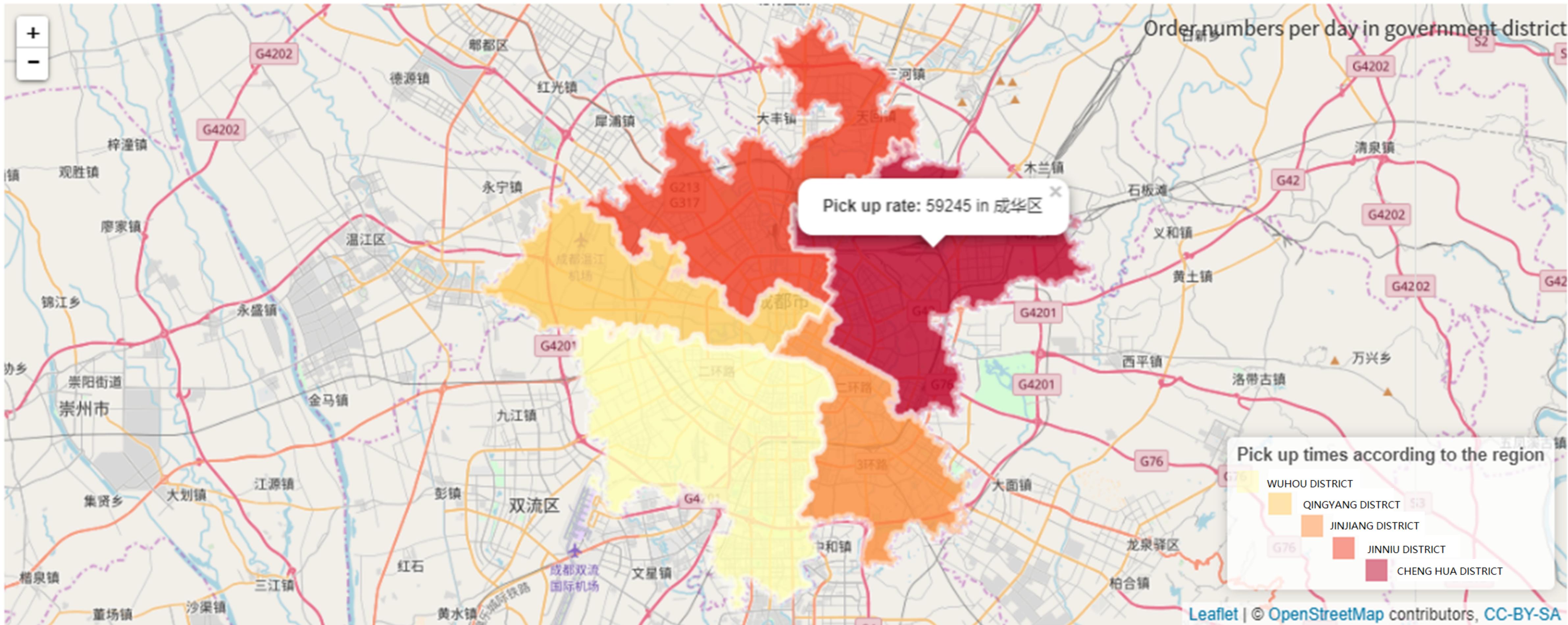
The overview of Didi market

The overview of Didi market according to Orders and Cars numbers

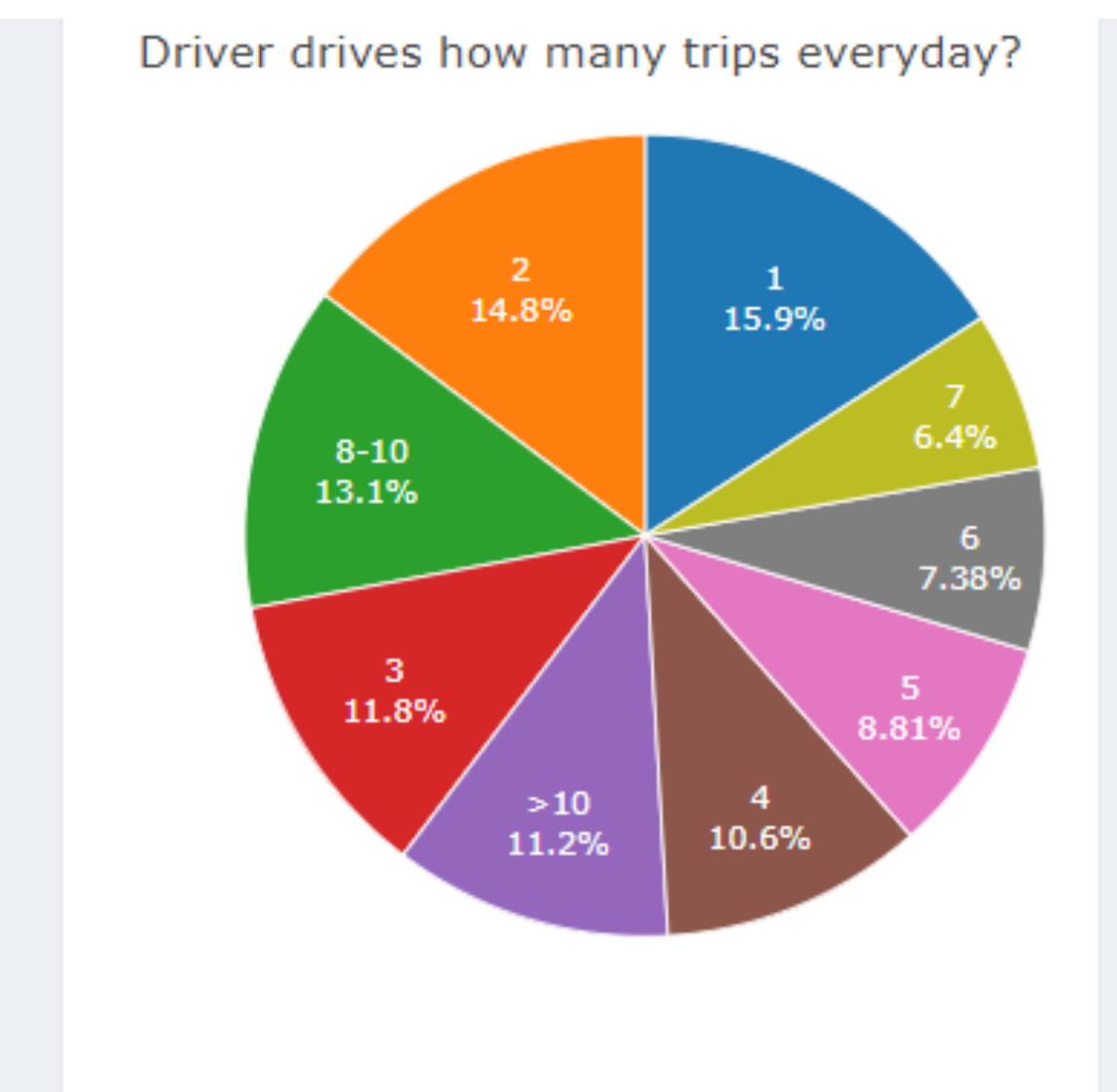
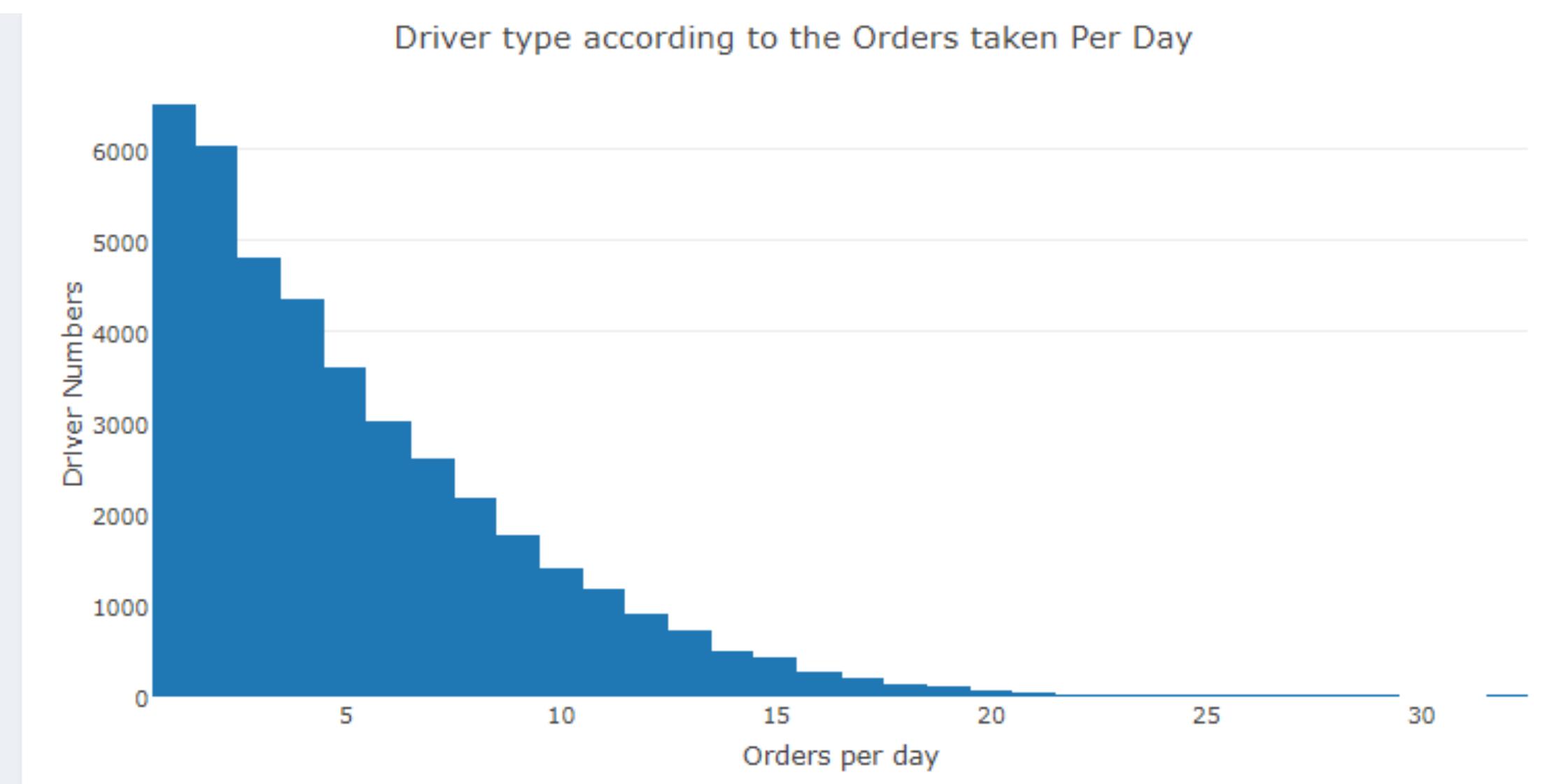
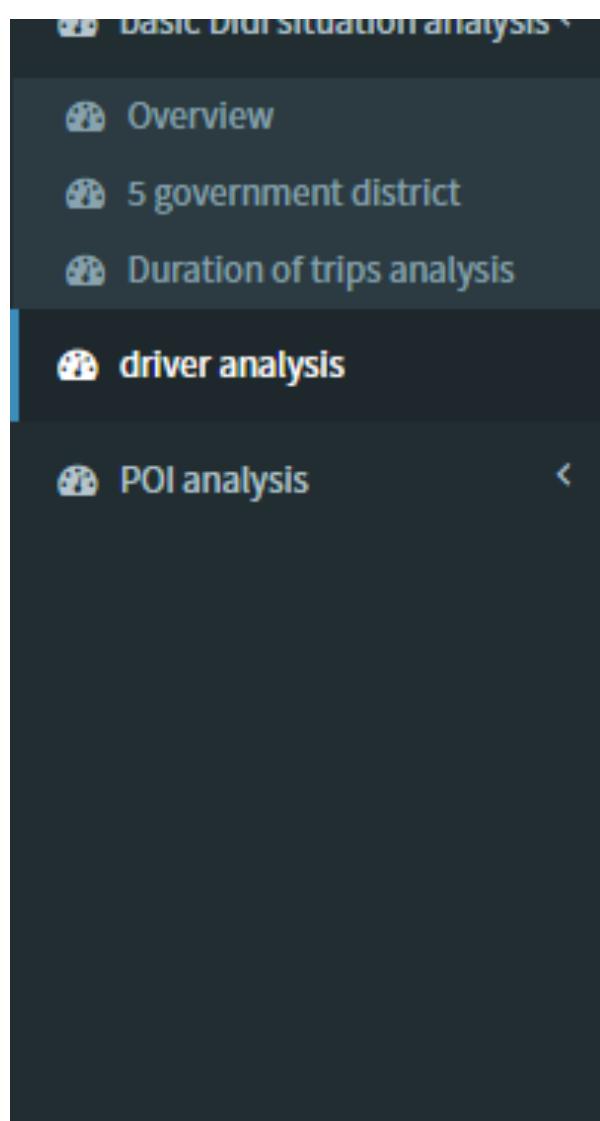
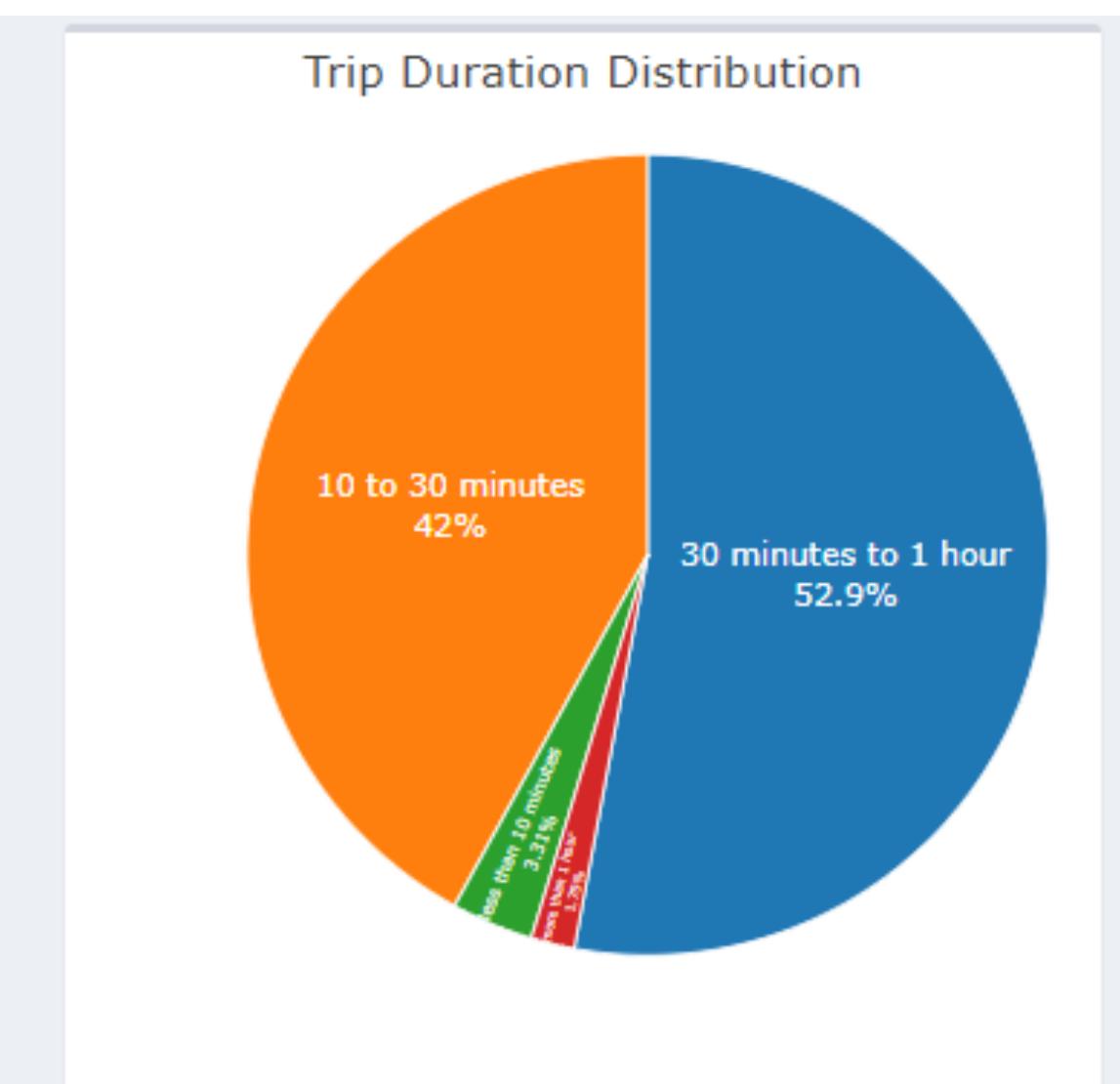
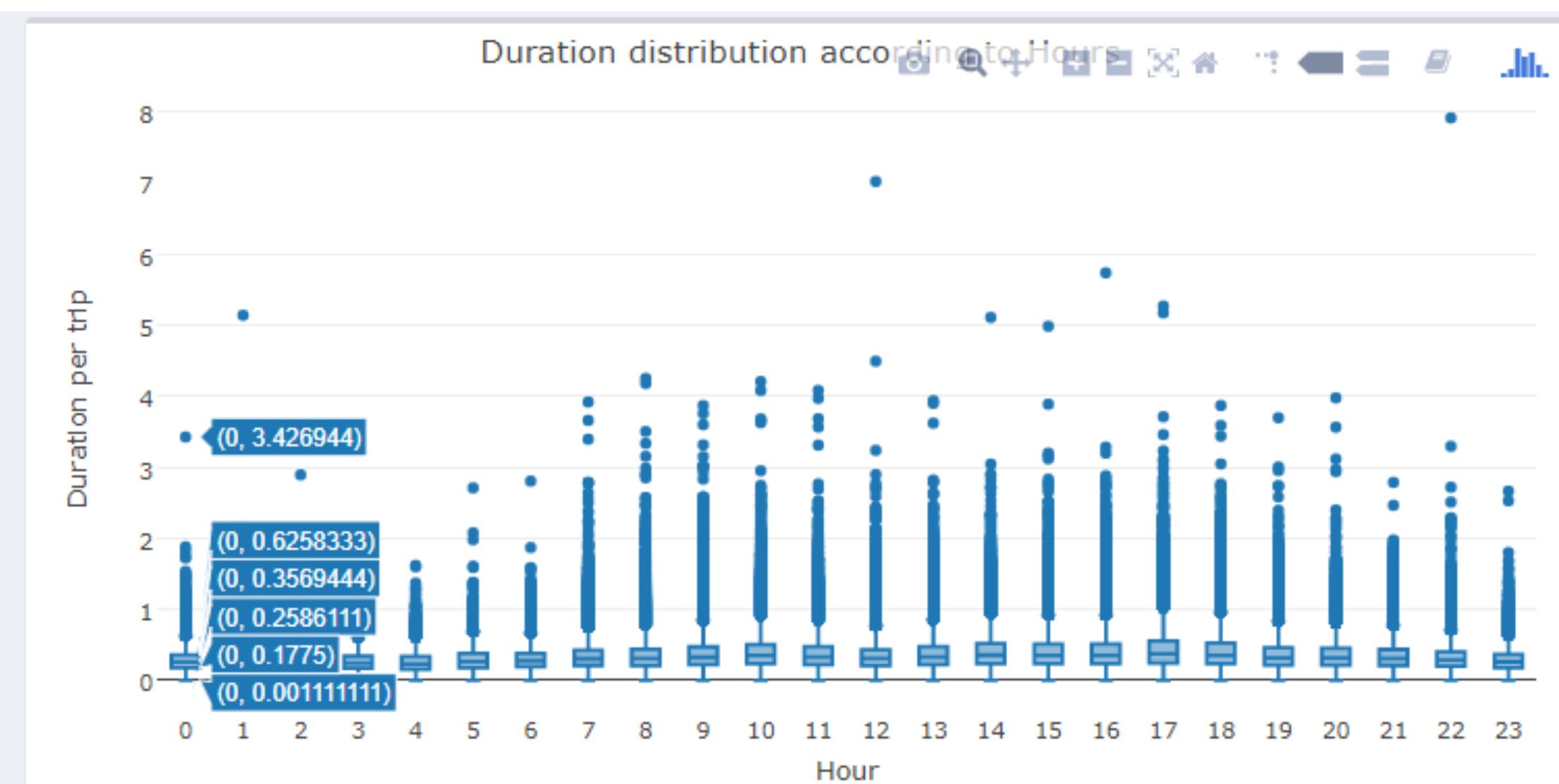
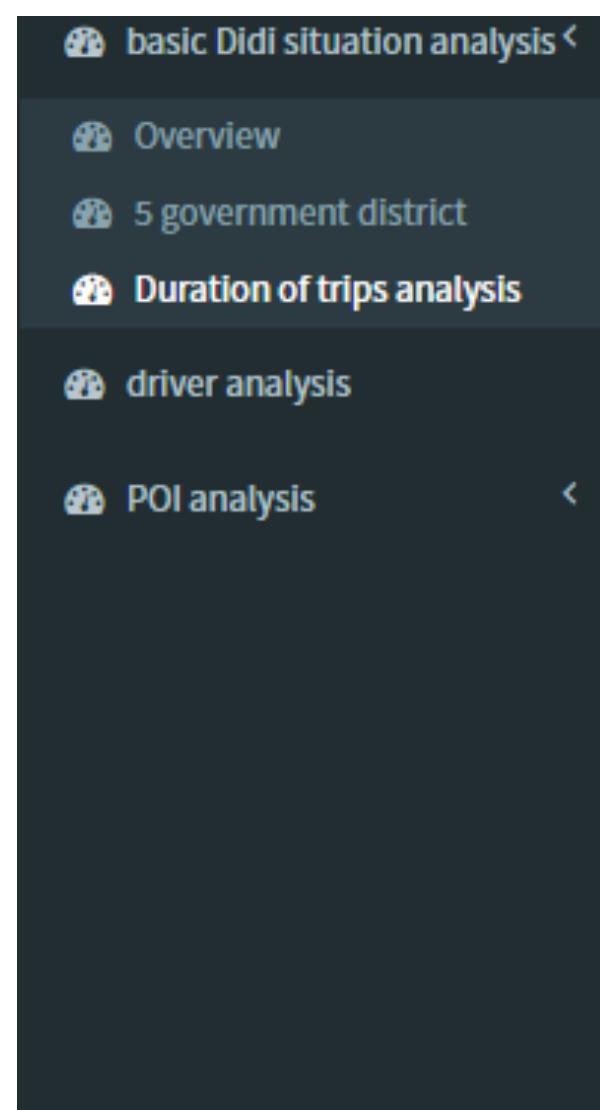
A grouped bar chart comparing the average number of orders (dark blue bars) and cars (light blue bars) per day. The Y-axis represents the number of units, ranging from 0 to 200k. The X-axis lists the days of the week: Sun, Mon, Tues, Wed, Thurs, Fri, Sat. The chart shows a general upward trend from Monday to Friday, followed by a slight decrease on Saturday.

Day	Number of orders	Number of cars
Sun	~200k	~40k
Mon	~195k	~38k
Tues	~205k	~38k
Wed	~210k	~38k
Thurs	~208k	~40k
Fri	~215k	~42k
Sat	~208k	~42k

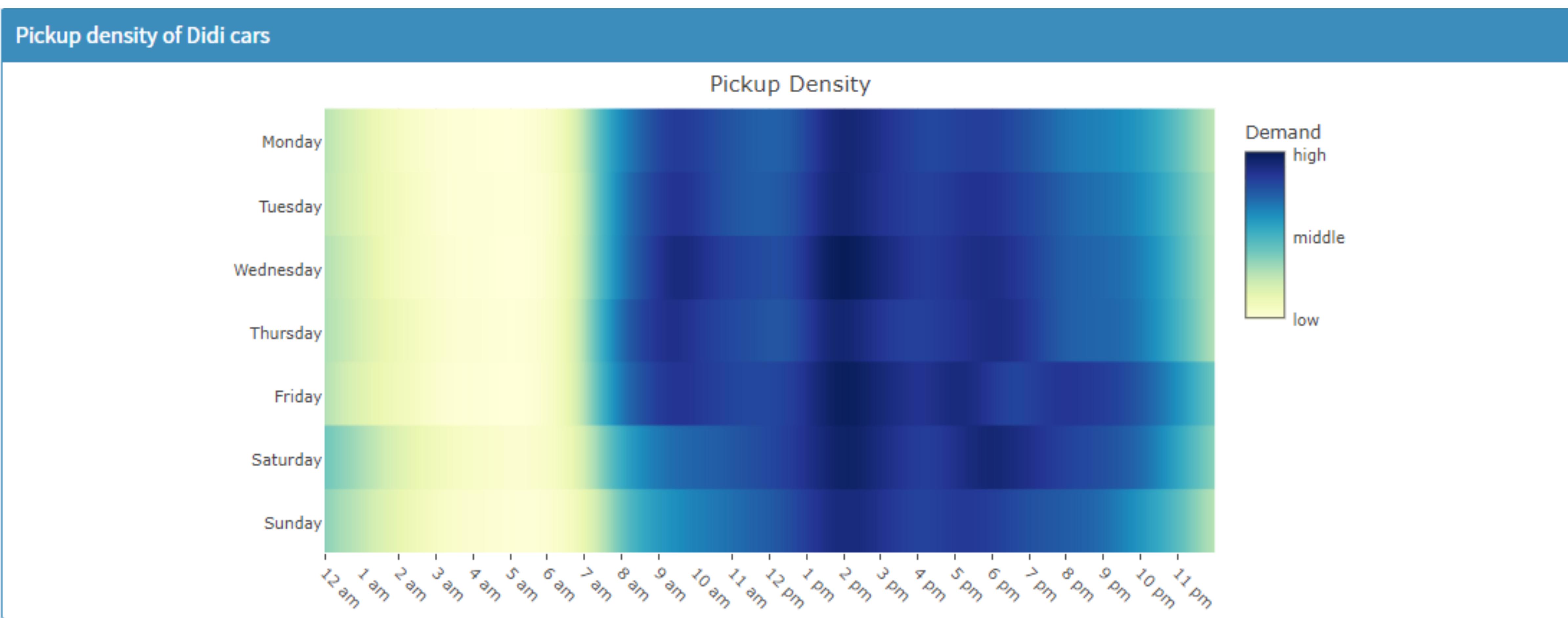
The Overview of DIDI in CHENGDU



The Overview of DIDI in CHENGDU



The Overview of DIDI in Chengdu





05

2nd Question:
How people use
DIDI(DEMO)



06

3rd Question:
COMPARATION
AND
IMPROVEMENT
(DEMO)

Thank you for your watching!

Q&A