



Chicago 2016 Taxi Rides Analysis

Group 5

Jingsi WU

Yidan WANG

Jingyi LIU

Xiaonan ZHAN

CONTENTS



Part 1

Database overview
& Initial analysis

Part 2

Creating the
relational data
model

Part 3

Analyze the data

Part 4

Additional
Analysis
And Conclusion



Part 1: Selecting

Chicago Taxi Rides January 2016

Why

Why this dataset?

Chicago residents and visitors took more than 27 million taxi rides, spend more than \$400 million.
Development of app-based rides services such as Uber and Lyft.

What

What is the dataset?

The database contains over 100 thousands taxi rides in Chicago for January 2016, which describe the rides from multiple aspects including companies, times, fares, duration, etc.

How

How to analyze?

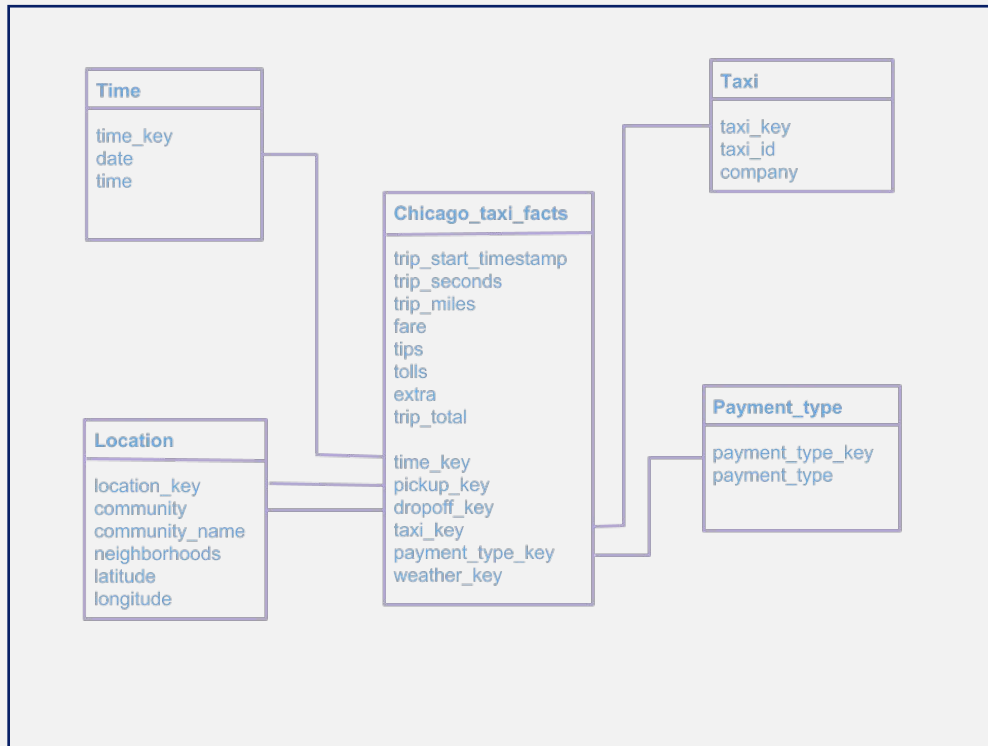
Explore how the behavior of Chicago taxi riders is being affected by location, time, etc.
Explore how taxi rides are priced and the elements which may affect the fares of rides.

id		taxi_id	trip_start_timestamp		time	trip_seconds	trip_miles	pickup_community_area		dropoff_community_area		fare
7615		251	2016-01-12 00:00:00		1:00:00AM	0.0	0.0	14		14		0.0
tips	tolls	extra	trip_total	payment_type	pickup_latitude	pickup_longitude	dropoff_latitude	dropoff_longitude	company	time_key		
0.0	0.0	0.0	0.0	Cash	41.968069	-87.721559063	41.968069	-87.721559063	KOAM Taxi Association	1849		

After
selected



Part 2: Wrangling



Create the
fact table

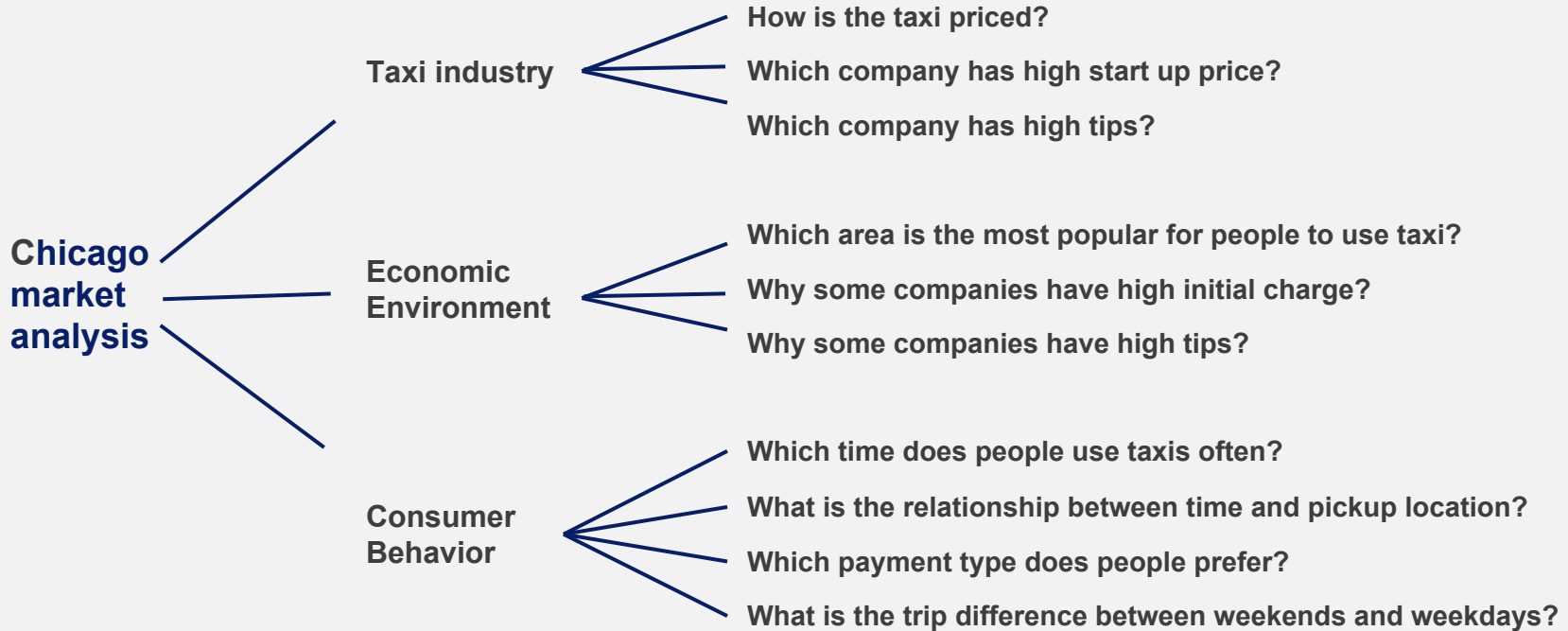
Clean the
Dimension table

Add Foreign
Key to fact table

Connect the fact
and the
dimension



Part 3: Analysis



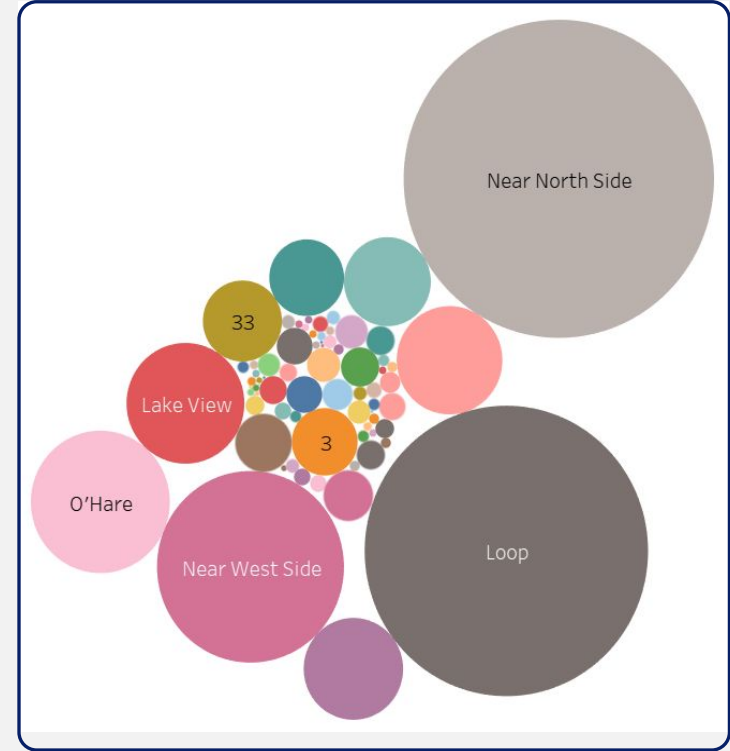
Part 3: Analysis

company	start_price
4053 - 40193 Adwar H. Nikola	43.11
3011 - 66308 JBL Cab Inc.	24.90
Suburban Dispatch LLC	22.96
6747 - Mueen Abdalla	12.89
1247 - 72807 Daniel Ayertey	11.65

company	average_tip
3669 - 85800 Jordan Taxi Inc	0.00
2092 - 61288 Sbeih company	0.09
6743 - 78771 Luhak Corp	0.14
5997 - 65283 AW Services Inc.	0.27
3094 - 24059 G.L.B. Cab Co	0.31

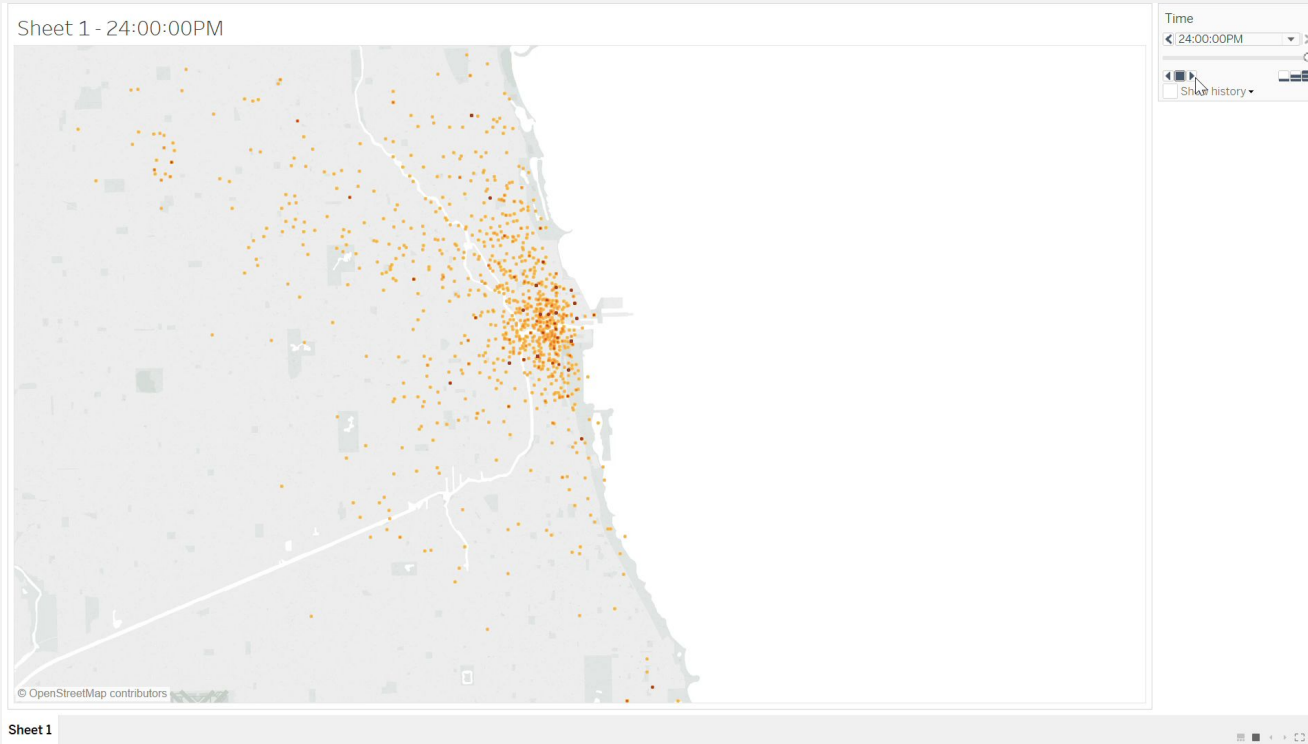
[71]:	company	community	start_price
	0694 - 59280 Chinesco Trans Inc	28	13.25
	0694 - 59280 Chinesco Trans Inc	6	3.25
	1085 - 72312 N and W Cab Co	28	3.25
	1085 - 72312 N and W Cab Co	8	8.50
	1247 - 72807 Daniel Ayertey	8	11.65
	2092 - 61288 Sbeih company	11	3.25
	2192 - 73487 Zeymane Corp	7	13.25
	2192 - 73487 Zeymane Corp	32	7.13
	2192 - 73487 Zeymane Corp	8	9.21
	2192 - 73487 Zeymane Corp	23	3.25

[70]:	company	community	average_tips
	0694 - 59280 Chinesco Trans Inc	28	0.00
	0694 - 59280 Chinesco Trans Inc	6	0.00
	1085 - 72312 N and W Cab Co	28	0.00
	1085 - 72312 N and W Cab Co	8	5.25
	1247 - 72807 Daniel Ayertey	8	2.40
	2092 - 61288 Sbeih company	11	0.00
	2192 - 73487 Zeymane Corp	7	0.00
	2192 - 73487 Zeymane Corp	32	0.50
	2192 - 73487 Zeymane Corp	8	0.31
	2192 - 73487 Zeymane Corp	23	0.00





Part 3: Analysis



Part 3: Analysis

Findings

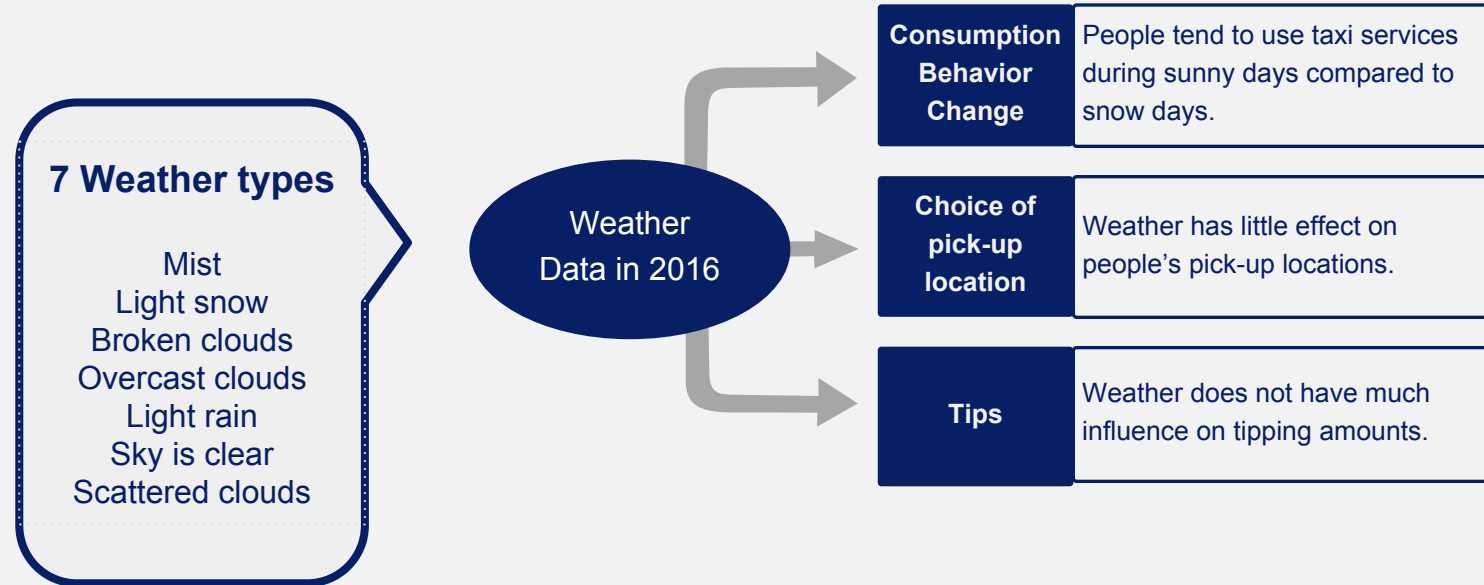
6:45-7:15 PM is the peak hour

People in Chicago prefer cash to pay the money

There is a periodic pattern. Every Friday reach a peak.

The place where people go after work is near their office.

Part 4: Additional Analysis and Conclusion



Conclusion

What we analyze?

When people tend to use taxis, which community area is popular destination, what is the pricing strategy, How weather influence consumer behavior

What tools we use?

Tableau, SQL, shells, pandas, Trifika, ect.

How Uber impact the industry?

The base price and price per mile of Uber are all cheaper than taxi.

What is the purpose of the analysis

Suggest Uber to maintain the competitive advantage, balance high demand at rush hours, optimizing the whole operation

Porter's 5 Forces

All the forces are quite high except the bargaining power of suppliers. So the whole industry is quite competitive

