Optimal Driver Positioning

Anthony Chan

Defining the Scope/Problem

- Stakeholders
  - City X (GM, drivers, and riders)
  - Uber
- Location
  - Chelsea Court (CC) neighborhood
- What is the Problem?
  - City demand
  - Surge pricing
  - ETA
- What is the expected solution?
  - Provide statistics to confirm or reject hypothesis

Initial Data Exploration

- Data cleaning
  - Conversion of timestamps
    - Adding columns for months, hours, and days
  - Acknowledgement of missing data (ETA, ATA)
- Table merging
  - Driver_trips and rider_trips
- Masking
  - Chelsea Court vs not Chelsea Court (start_geo only)
  - Trip status: completed vs not completed

Basic Statistics - All Trips

	count	mean	std	min	25%	50%	75%	max
actual_time_to_arrival	**53583.0**	**2.675140**	**2.585005**	**0.02**	**0.800**	**1.970**	**3.8200**	**70.320**
surge_multiplier	**60000.0**	**1.158863**	**0.335814**	**1.00**	**1.000**	**1.000**	**1.2000**	**4.800**
driver payout	59854.0	6.096275	4.817392	0.00	4.240	4.940	7.0600	175.870
estimated_time_to_arrival	**57499.0**	**3.270866**	**2.148824**	**0.02**	**1.730**	**2.830**	**4.3300**	**53.420**
trip_price_pre_discount	60000.0	7.601100	6.026331	0.00	5.296	6.168	8.8080	219.832
rider_payment	60000.0	7.551018	6.002798	0.00	5.296	6.120	8.7048	219.832

Location Breakdown

	Start_geo	%	End_geo	%
Allen Abby	10493	17.5	10765	18.0
Blair Bend	4372	7.3	4606	7.7
Chelsea Court	**40363**	**67.3**	**39828**	**66.4**
Daisy Drive	4772	8.0	4801	8.0

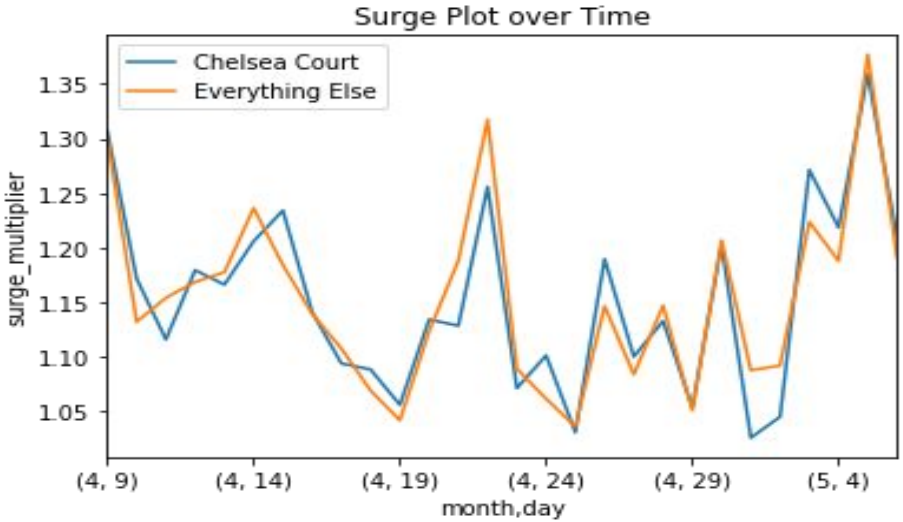
Trip Status

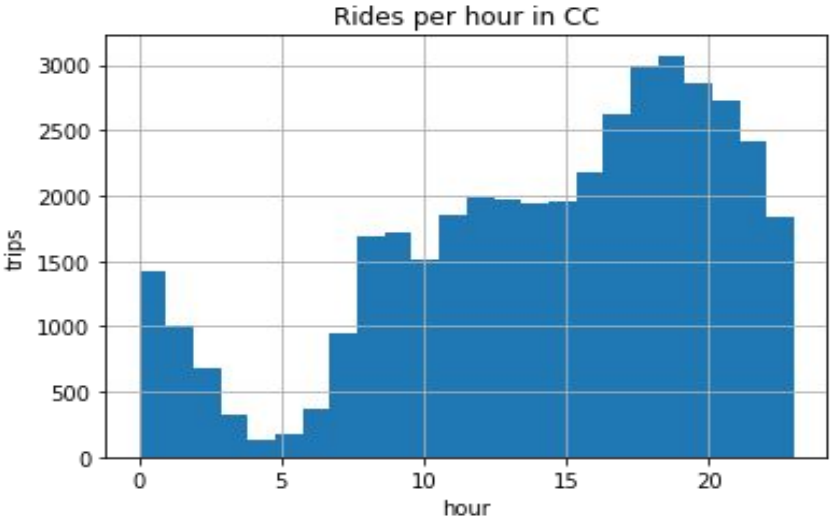
trip_status	Overall	CC
completed	**54139**	**36286**
failed	49	35
rider_canceled	5777	4025
unfulfilled	35	17

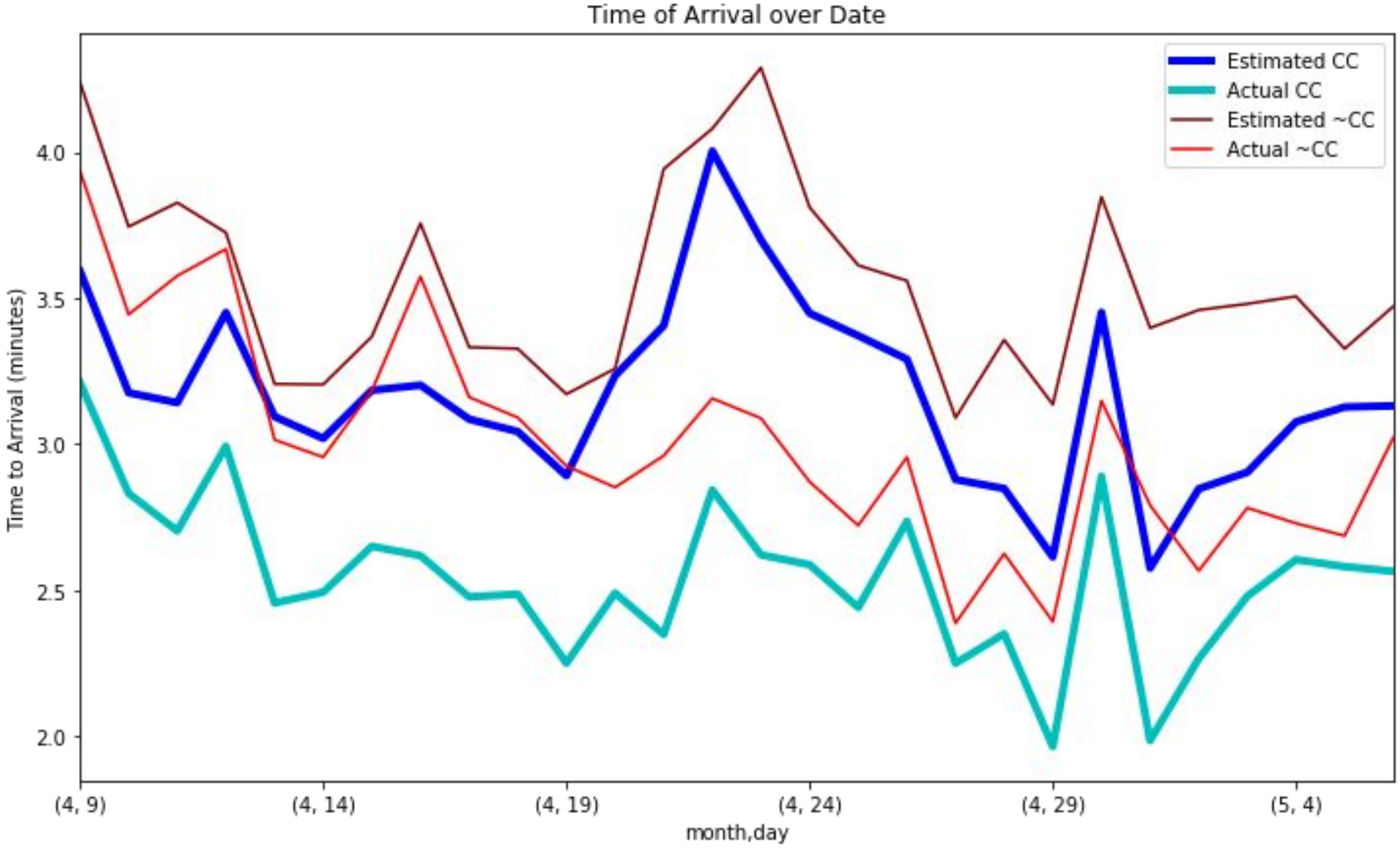
Key Metrics

- From City X GM
  - Price Surge
  - Estimated time to arrival (ETA)
  - Starting Location (CC)
- Additional Analysis
  - Actual time to arrival (ATA)
  - Trip status
- Engineered Metrics
  - Trips / requests ratio
  - Month, hour, day breakdown

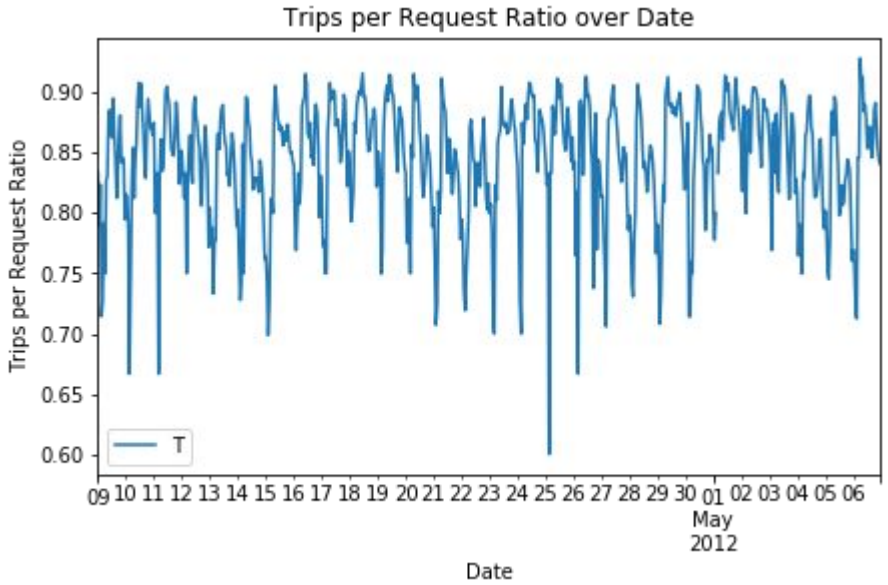
Price Surge over Time

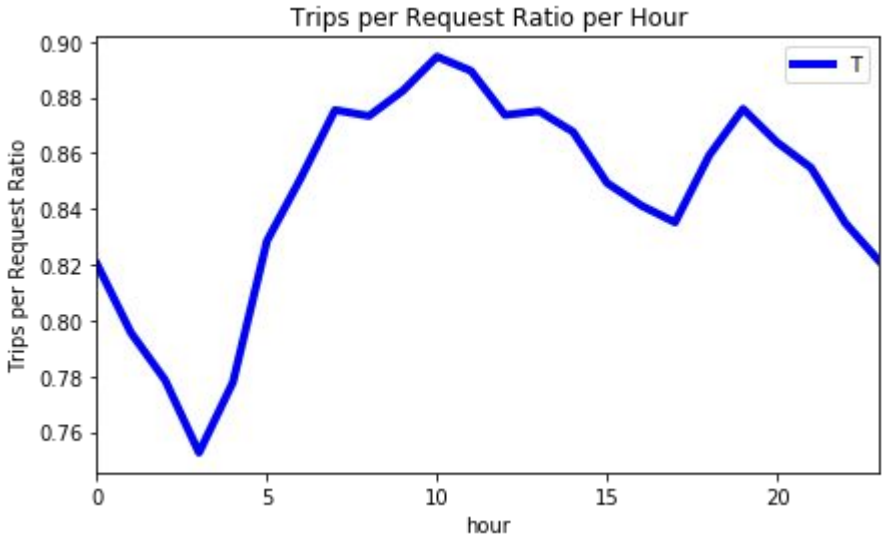





Time of Arrival over Date
Estimated CC
Actual CC
Estimated ~CC
Actual ~CC
Time to Arrival (minutes)
4.0
3.5
3.0
2.5
2.0
(4, 9)
(4, 14)
(4, 19)
(4, 24)
(4, 29)
(5, 4)
month,day

Ride Demand over Time

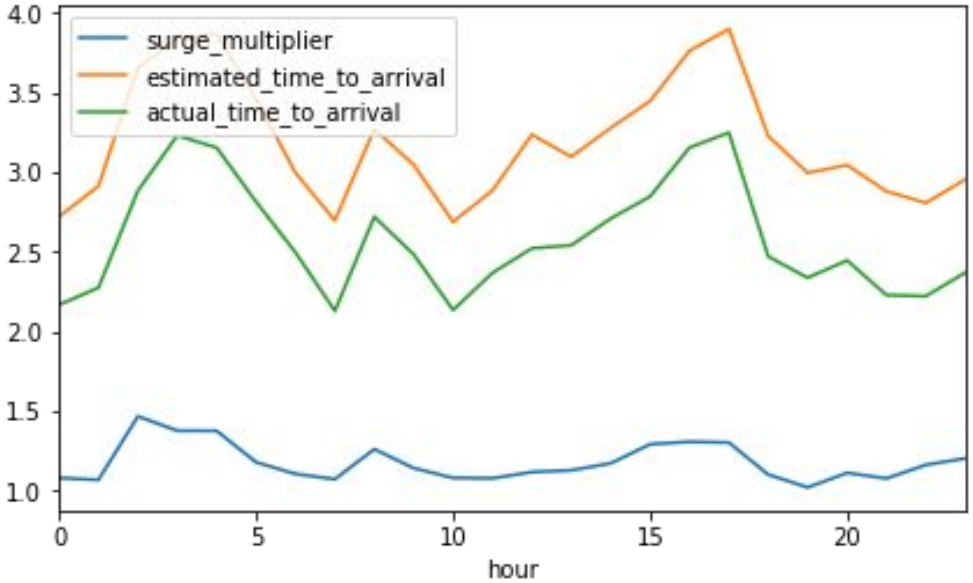




Q1: Data Driven View on Potential Issue

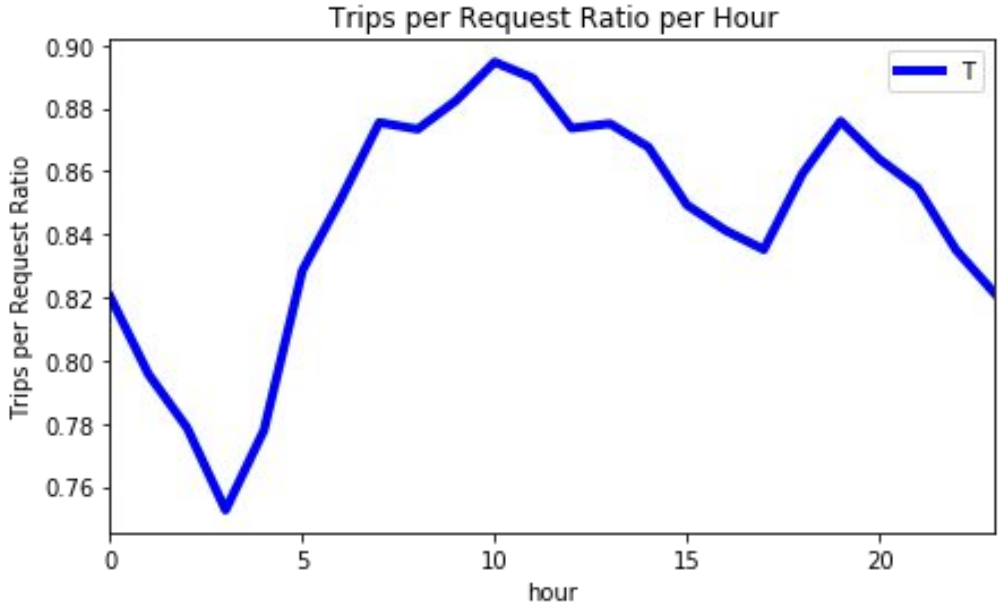
- Q1a: Disproportionate levels of surge pricing and higher ETA in CC during rush hour (defined as hour 15-19)
- A1a: Based on our data, average surge and ETA are both lower in CC than outside.
- Additional calculations show a rush hour Surge = 1.20, ETA = 3.46, ATA = 2.81, and non-rush hour Surge = 1.17, ETA = 3.12, ATA 2.52
- With Price Surging greater by 2.5%, and ETA by ⅓ of a minute, we reject the claims of disproportionate levels of price surging, and 1-2 minutes higher ETA during rush hour.

Location	Surge	ETA	ATA
CC	1.151	3.01	2.57
Not CC	1.157	3.39	3.00



Q1: Data Driven View on Potential Issue

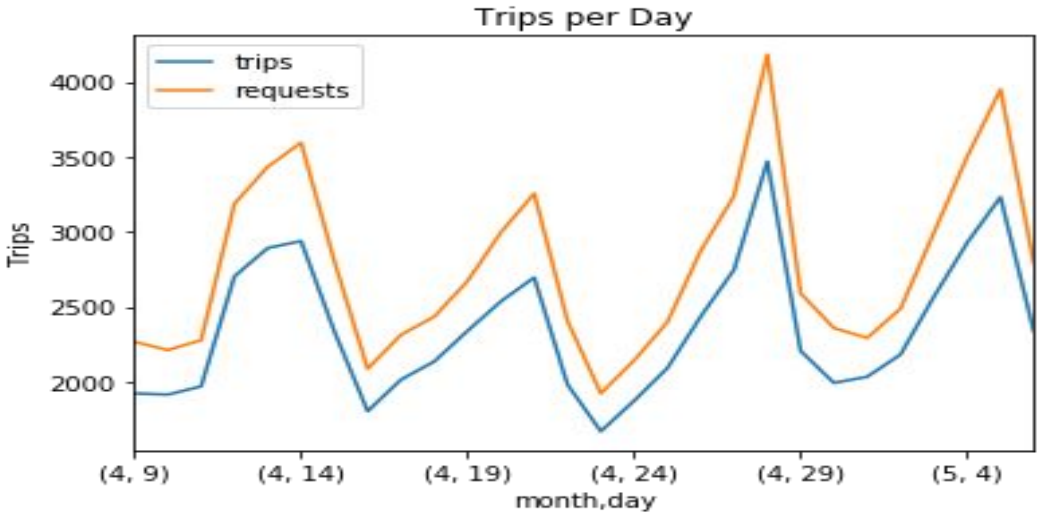
- Q1b: Do a better job of letting drivers know about demand in CC during rush hour
- A1b: Based on afternoon rush hours between 15 and 19, ratio of trips per request drops. This means Uber will need to address this issue by moving more drivers to high demand locations in CC.



Q1: Data Driven View on Potential Issue

- Q1c: Inconsistent growth in April and May
- A1c: Based on our data, we do see marginal growth through the month of April. However, it could be due to an event spike in late April. In order to give a more accurate assessment on growth, a longer time duration and more data will need to be collected and evaluated.

	CC	Not CC
Completed	36286	17853
Rider Cancelled	4025	1752
Failed	35	14
Unfulfilled	17	18



Q2: A/B Testing - Rush Hour Demand

No Trips = Requests - Trips

Control is defined as non rush hours: 0-14, 20-23

Test is defined as rush hours: 15-19

Null hypothesis: no difference between trips and no trips during rush hour

Chi squared test: T-statistic = 0.268, p-value = 0.87 > 0.05: accept null hypothesis

Observed	Control	Test	Total		Expected	Control	Test	Total
Trips	45209	20727	65936		Trips	45233	20703	65936
No Trips	8063	3655	11718		No Trips	8039	3679	11718
Total	53272	24382	77654		Total	53272	24382	77654

Recommend Actionable Next Steps and Case Study

- Length of study
  - Include several months, keeping in mind seasonal changes and major events
- Additional data collection
  - distance traveled to pick up
  - Distance traveled during drive
- Temporary solutions
  - Provide additional vehicles and incentive during high demand hours and events
- Demand - Expected vs Actual
  - Arma model
- improvements on metrics
  - Moving vs idling time (Left turns)
- Other
  - Supply hours vs demand
  - Location accuracy
  - Confusion of pick up location

Edge Cases

High surge values for certain days:

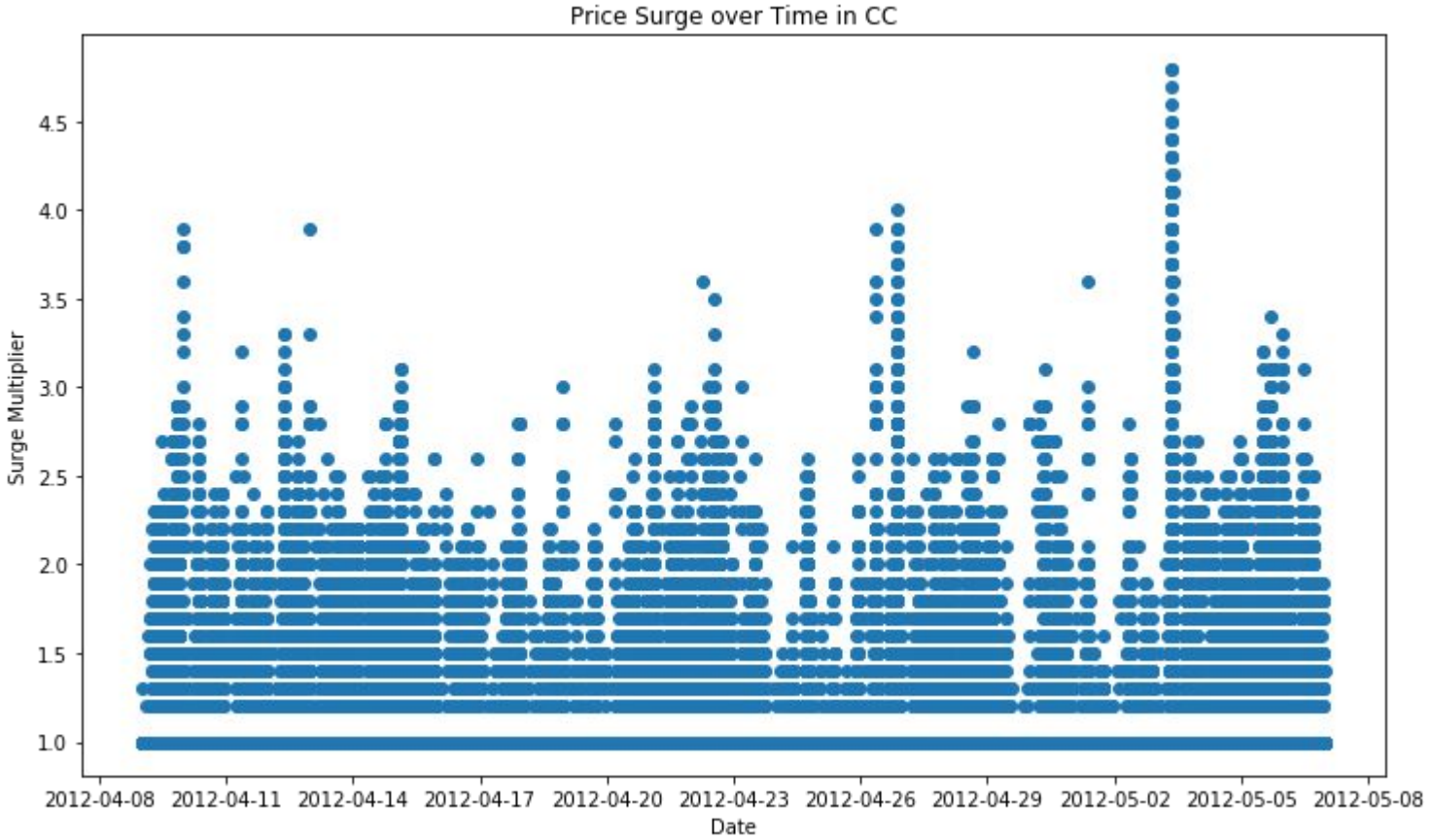
4/9/2012: 1.30

4/25/2012: 1.03

4/26/2012: 1.17

5/3/2012: 1.26

Average: 1.15



Summary

1. Metrics - rush hour
    a. Disproportionate levels of surge pricing: rejected
    b. Higher ETA (1-2 minutes) : rejected
    c. More Driver Demand: accepted
2. A/B testing - rush hour
    a. Null hypothesis accepted
3. Inconsistent growth in April and May
    a. Need more information, longer study

Methods of Analysis

- Languages
  - Python
  - https://github.com/antawnchan/Uber
- Database
  - Pandas
- Data Visualization
  - Matplotlib
- Basic Calculations
  - Excel

Questions?