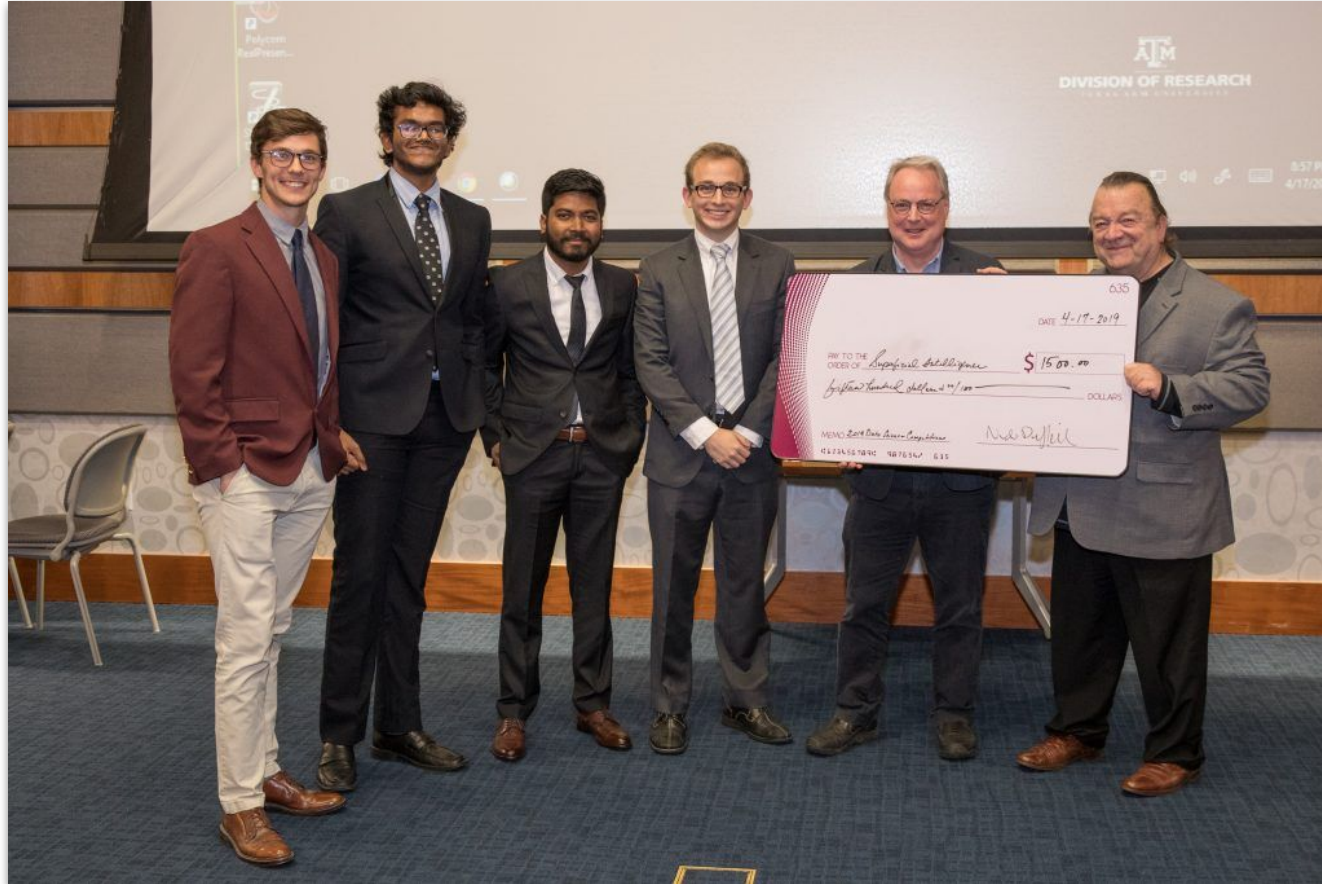




LA Metro Bike Share

"Analyze data taken from more than 600,000 bicycle commuter trips made in Los Angeles and **develop business solutions** for LA Bike Share, a for-profit company in Los Angeles."





1st Place in Grad Texas A&M 2019

Collaborators:

- Josiah Coad
- Chinmay Phulse
- Brandon Walker
- Sheelabhadra Dey

April 2019



2nd Place in Grad Texas A&M 2018

Collaborators:

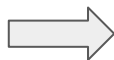
- Josiah Coad
- Erica
- Savinay

April 2018

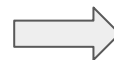
Research Question



Data Collection



Data Cleaning

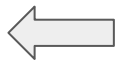


Exploration &
Visualization

Data Science



Update Domain
Knowledge



Interpretation of
Result



Post Hoc Analysis



Model Fitting
"Machine Learning"

"Where should we place new bike share stations in LA city?"

- How successful are current stations? Why?
- Where do people want new stations? Why?

Notes

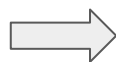


- Think outside the box!
- What questions aren't being asked?
- What are people wasting time on that could be automated?

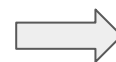
Research Question



Data Collection



Data Cleaning



Exploration &
Visualization

Data Science



Update Domain
Knowledge



Interpretation of
Result



Post Hoc Analysis



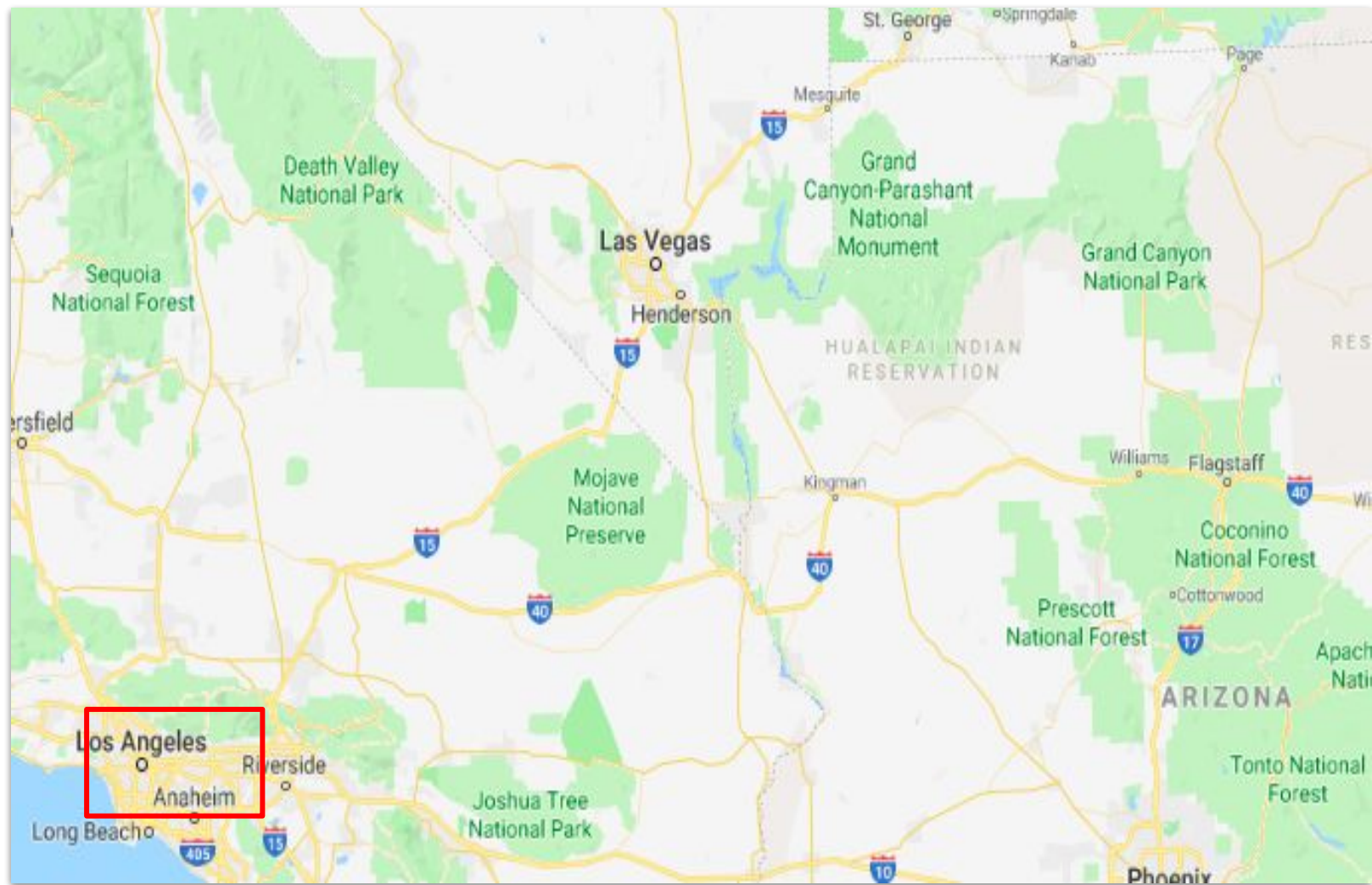
Model Fitting
"Machine Learning"

- Ride data (trip time and location)
- External data (census data)
- Web-scraped data (forums)

Notes



- Think outside the box!
- Start collecting data now.
- Streamline the data collection process for more efficient use.



Research Question



Data Collection



Data Cleaning

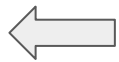


Exploration &
Visualization

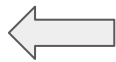
Data Science



Update Domain
Knowledge



Interpretation of
Result

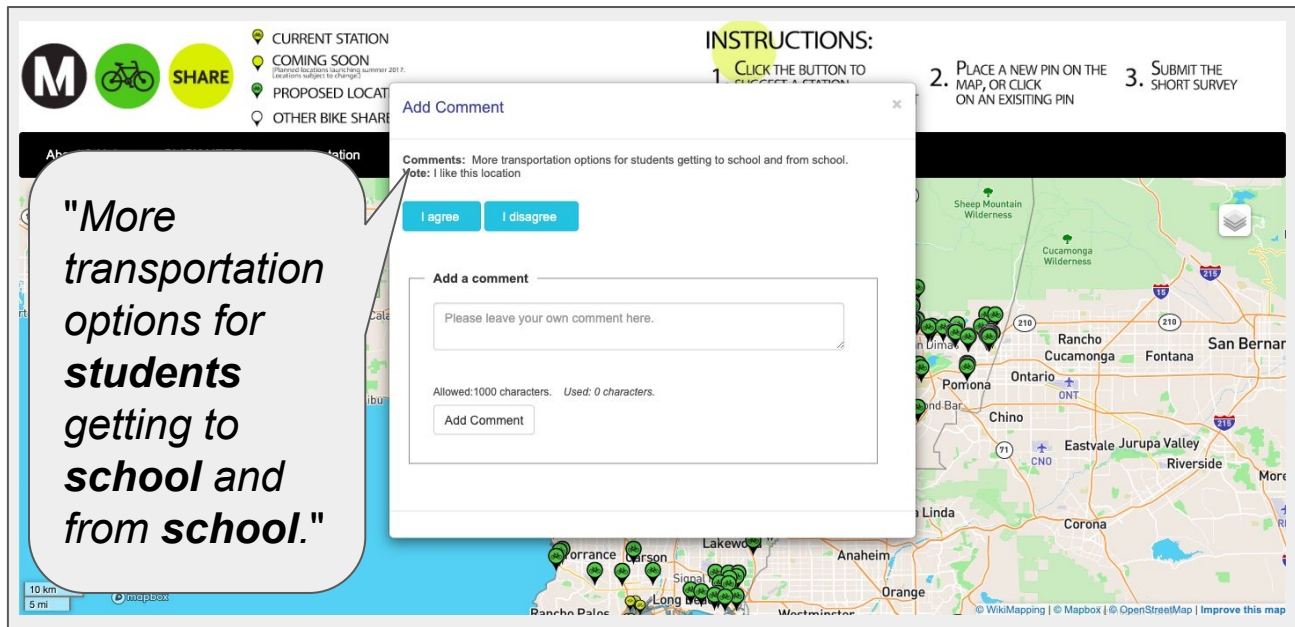


Post Hoc Analysis



Model Fitting
"Machine Learning"

WEB SCRAPED DATA



INSTRUCTIONS:

1. CLICK THE BUTTON TO SUCCESS A STATION
2. PLACE A NEW PIN ON THE MAP, OR CLICK ON AN EXISTING PIN
3. SUBMIT THE SHORT SURVEY

Comments: More transportation options for students getting to school and from school.
Vote: I like this location

Add Comment

Add a comment

Please leave your own comment here.

Allowed: 1000 characters. Used: 0 characters.

Add Comment

"More transportation options for students getting to school and from school."

Notes



- Think outside the box!
- Start collecting data now.
- Streamline the data collection process for more efficient use.

Research Question



Data Collection



Data Cleaning

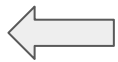


Exploration &
Visualization

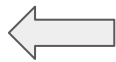
Data Science



Update Domain
Knowledge



Interpretation of
Result



Post Hoc Analysis



Model Fitting
"Machine Learning"

- How to fill in missing values?
- How to format data?
- Feature engineering
- *Python, Jupyter and Pandas*

Notes



- This is time consuming! Try to get data as clean as possible from the beginning.
- Bias introduced when cleaning. Document this!

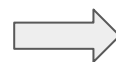
Research Question



Data Collection



Data Cleaning

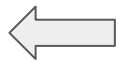


Exploration &
Visualization

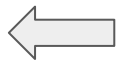
Data Science



Update Domain
Knowledge



Interpretation of
Result



Post Hoc Analysis



Model Fitting
"Machine Learning"

WEB SCRAPED DATA

comment = *"More transportation options for **students** getting to **school** and from **school**."*



extract_topics(comment)



students, school



| comment | timestamp | location_lat | location_lgn | topics |
|---------|-----------|--------------|--------------|-------------------|
| More... | 2343243.2 | 118.2437 | 34.0522 | [student, school] |

Notes



- This is time consuming! Try to get data as clean as possible from the beginning.

- Bias introduced when cleaning. Document this!

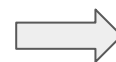
Research Question



Data Collection



Data Cleaning

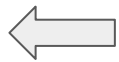


Exploration & Visualization

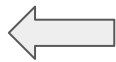
Data Science



Update Domain Knowledge



Interpretation of Result



Post Hoc Analysis

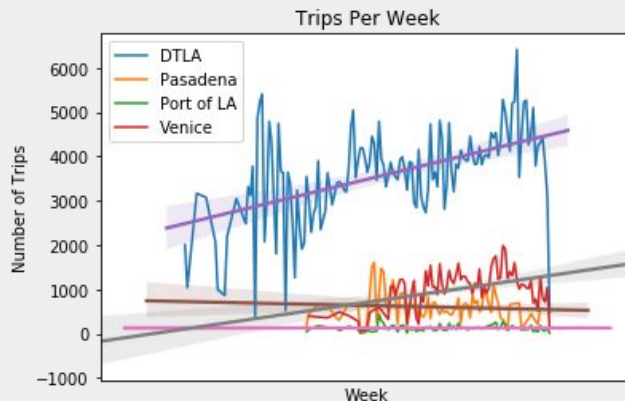
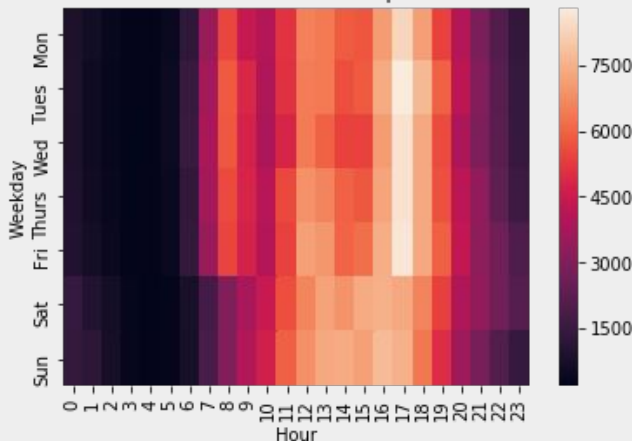


Model Fitting
"Machine Learning"

TRIP DATA

Our riders are commuters.

Total Number of Trips



Downtown LA usage
is on the rise!

Notes



- Visualizing is the most important part of the process.
- Convey a message.
- Test your assumptions.

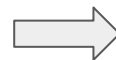
Research Question



Data Collection



Data Cleaning

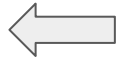


Exploration & Visualization

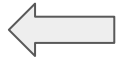
Data Science



Update Domain Knowledge



Interpretation of Result

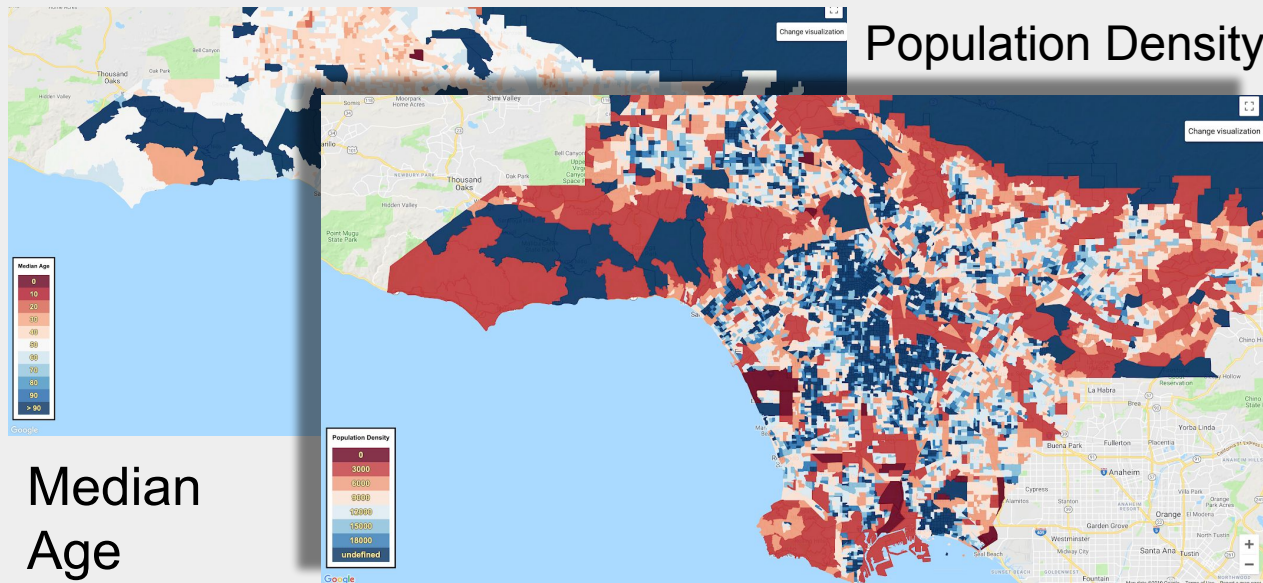


Post Hoc Analysis



Model Fitting
"Machine Learning"

CENSUS DATA



Notes



- Visualizing is the most important part of the process.
- Convey a message.
- Test your assumptions.

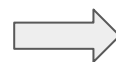
Research Question



Data Collection



Data Cleaning

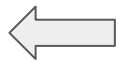


Exploration &
Visualization

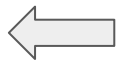
Data Science



Update Domain
Knowledge



Interpretation of
Result



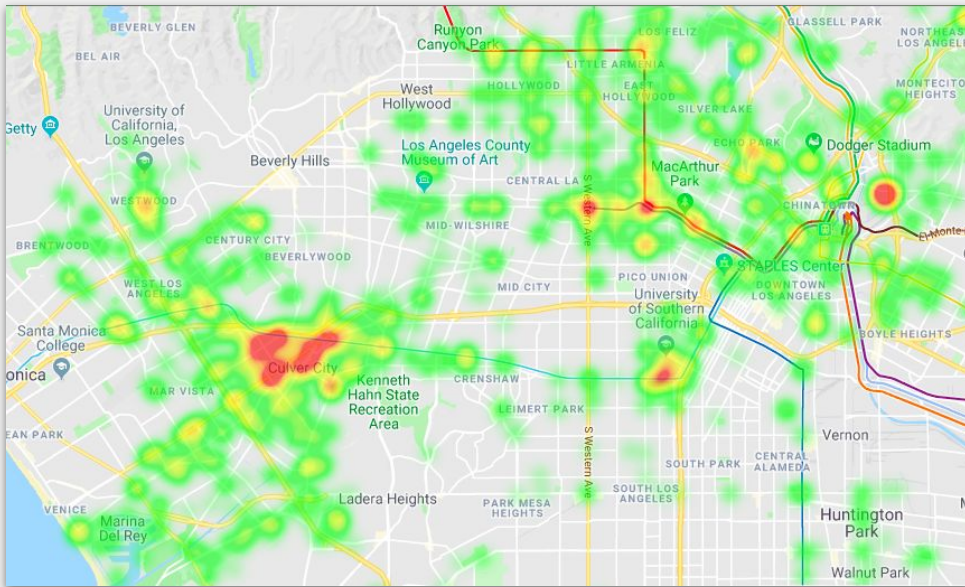
Post Hoc Analysis



Model Fitting
"Machine Learning"

WEB SCRAPED DATA

Heatmap
of LA
Bikeshare
Comment
Activity



Notes



- Visualizing is the most important part of the process.
- Convey a message.
- Test your assumptions.

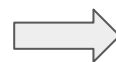
Research Question



Data Collection



Data Cleaning

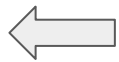


Exploration &
Visualization

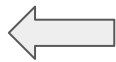
Data Science



Update Domain
Knowledge



Interpretation of
Result

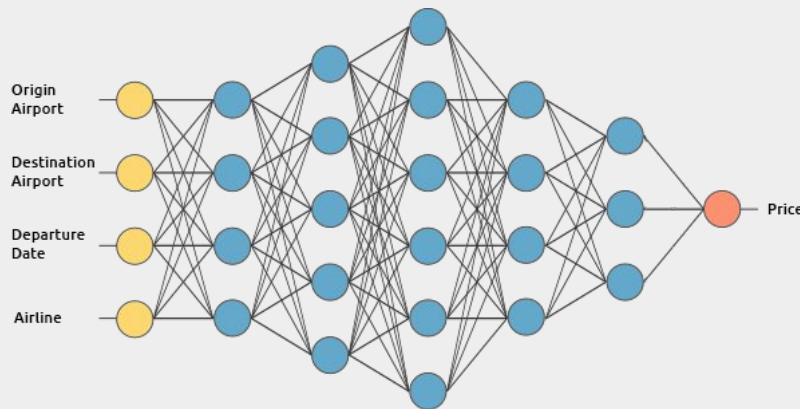
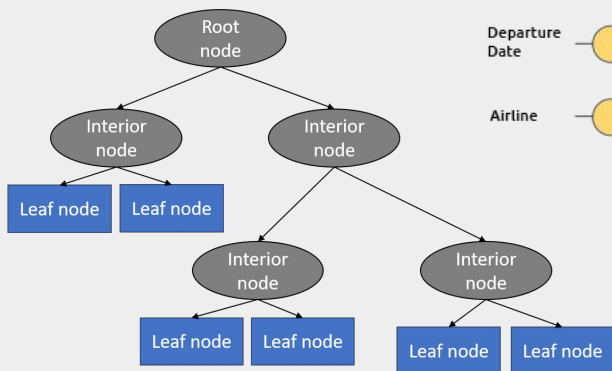


Post Hoc Analysis



Model Fitting
"Machine Learning"

Decision Trees



Neural Networks

Notes



- So many types of ML. Focus on how to pick the right one.
- The theory here is heavy. Software packages that do this for you! *sklearn*

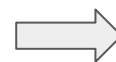
Research Question



Data Collection



Data Cleaning

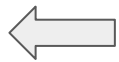


Exploration &
Visualization

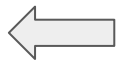
Data Science



Update Domain
Knowledge



Interpretation of
Result



Post Hoc Analysis



Model Fitting
"Machine Learning"

Markov Chain Modeling

| Station ID | Percentage of Total Bikes | Expected Number of Bikes |
|------------|---------------------------|--------------------------|
| 4215 | 0.03034476 | 37.63 |
| 4214 | 0.02443431 | 30.3 |
| 4210 | 0.02258008 | 28 |
| 3005 | 0.0206907 | 25.66 |
| 4207 | 0.0087878 | 10.9 |
| 4125 | 0.00163297 | 2.02 |
| 4127 | 0.00221579 | 2.75 |
| 4135 | 0.00343244 | 4.5 |

Stations overloaded

Stations underloaded

Notes

- So many types of ML. Focus on how to pick the right one.
- The theory here is heavy. Software packages that do this for you! *sklearn*

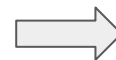
Research Question



Data Collection



Data Cleaning

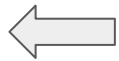


Exploration &
Visualization

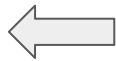
Data Science



Update Domain
Knowledge



Interpretation of
Result



Post Hoc Analysis

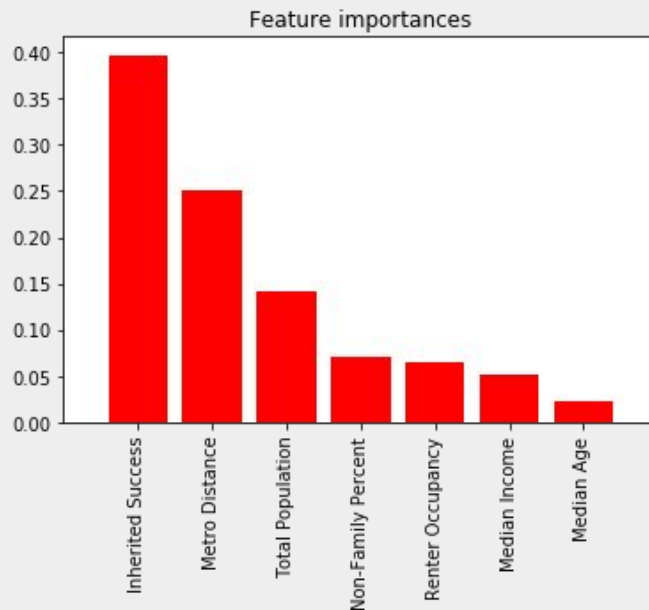


Model Fitting
"Machine Learning"

Feature Importance

Predict: **Station Popularity**
Using (in order of imp.):

1. Number of rides in surrounding stations
2. Metro distance
3. Population Density
4. Percent Ppl Single
5. Percent Ppl Renting
6. Median Age



Notes

- Feature importance
- Results of prediction, forecasting
- Report performance of machine learning systems

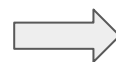
Research Question



Data Collection



Data Cleaning

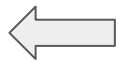


Exploration &
Visualization

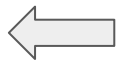
Data Science



Update Domain
Knowledge



Interpretation of
Result

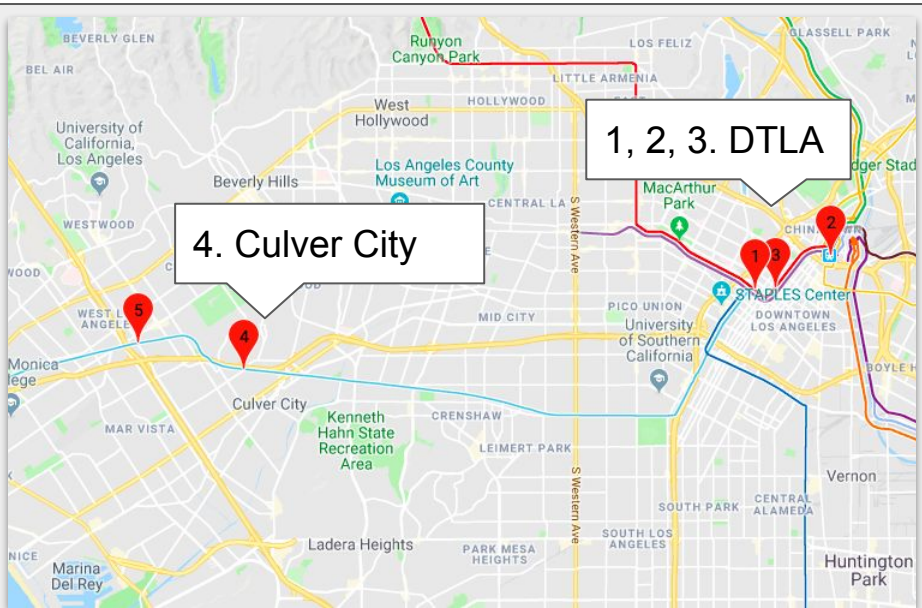


Post Hoc Analysis



Model Fitting
"Machine Learning"

Station Placement Using Machine Learning



Notes



- Feature importance
- Results of prediction, forecasting
- Report performance of machine learning systems

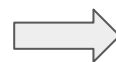
Research Question



Data Collection



Data Cleaning

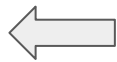


Exploration &
Visualization

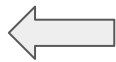
Data Science



Update Domain
Knowledge



Interpretation of
Result

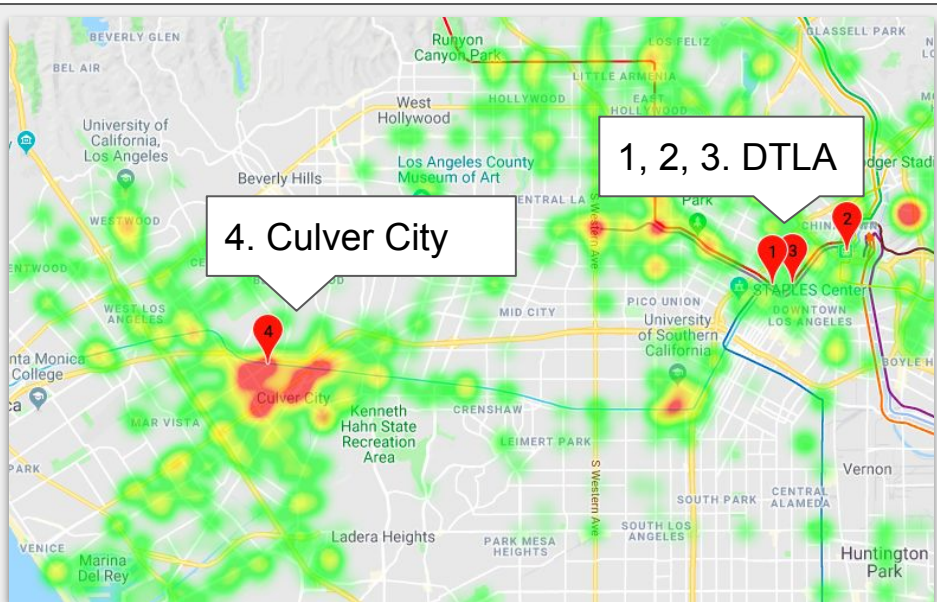


Post Hoc Analysis



Model Fitting
"Machine Learning"

Station Placement Using Machine Learning



Notes



- Feature importance
- Results of prediction, forecasting
- Report performance of machine learning systems

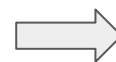
Research Question



Data Collection



Data Cleaning

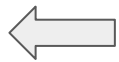


Exploration &
Visualization

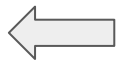
Data Science



Update Domain
Knowledge



Interpretation of
Result



Post Hoc Analysis



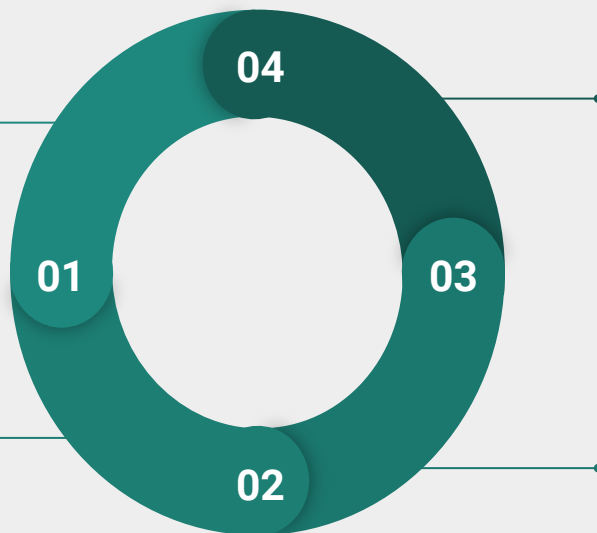
Model Fitting
"Machine Learning"

Proximity

Metro Expo Line,
Koreatown,
UCLA, Bus Stops
and Bike Paths

DTLA

Downtown LA
is popping.
Put more
stations here.



User Base

Young
Single
Commuting

Culver City

Your next big
move should
be Culver City

Notes



- Answer "So what?"
- Make it actionable
- All your conclusions must be supported by your data analysis

Tools if Using Python

- Code environment:
 - Jupyter notebook (google colab)
- Data Wrangling
 - Pandas
- Data Viz
 - High Level: Tableau/Data Studio
 - Mid Level: Plotly, Seaborn
 - Low Level: Matplotlib
- Machine Learning
 - Sklearn
- Presenting Options
 - (Static) Slides
 - (Dynamic) Plots on Colab
 - Interactive Website (bonus points)