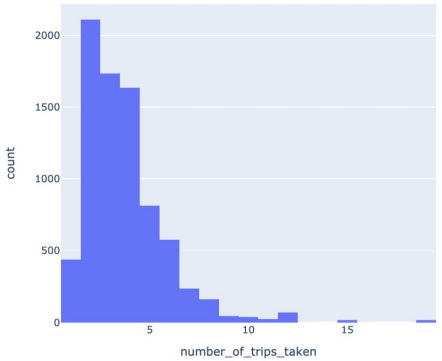
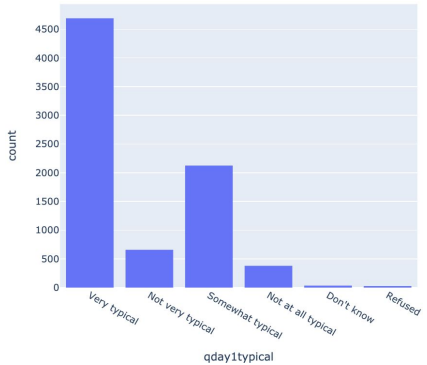
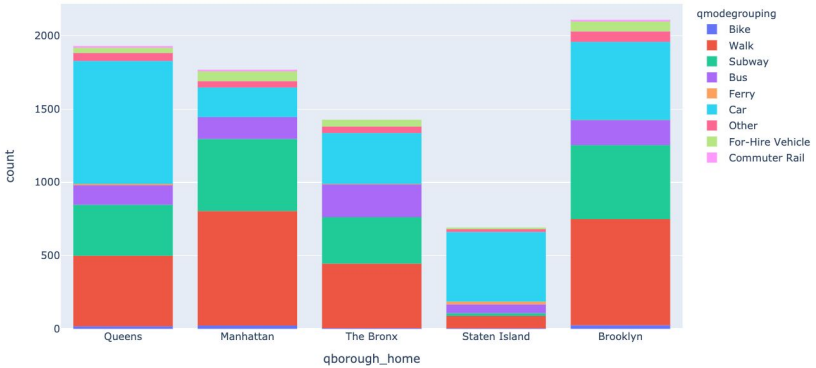

# Alternative Transportation Options For Trips in NYC

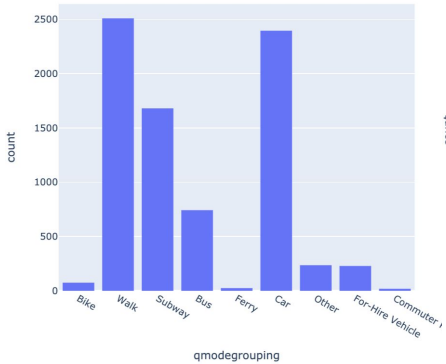
Brittany Fowle
Flatiron Data Science 2020

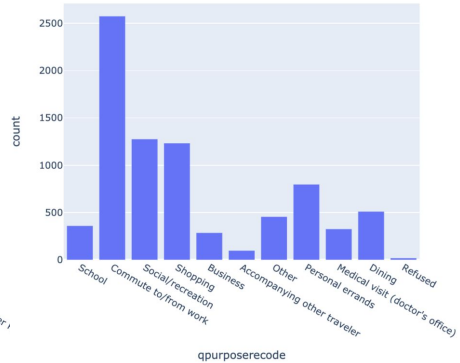
# EDA

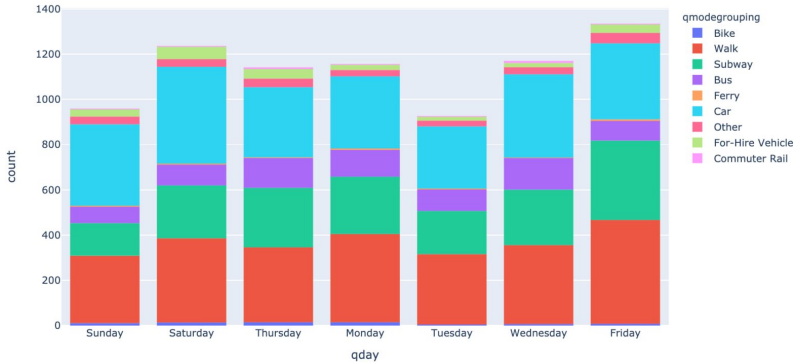

count
number_of_trips_taken


count
Very typical
Not very typical
Somewhat typical
Not at all typical
Don't know
Refused
qday1typical


qmodegrouping
Bike
Walk
Subway
Bus
Ferry
Car
Other
For-Hire Vehicle
Commuter Rail
count
Queens
Manhattan
The Bronx
Staten Island
Brooklyn
qborough_home


count
Bike
Walk
Subway
Bus
Ferry
Car
Other
For-Hire Vehicle
qmodegrouping


count
School
Commute to/from work
Social/recreation
Shopping
Business
Accompanying other traveler
Other
Personal errands
Medical visit (doctor's office)
Dining
Refused
qpurposerecode


qmodegrouping
Bike
Walk
Subway
Bus
Ferry
Car
Other
For-Hire Vehicle
Commuter Rail
count
Sunday
Saturday
Thursday
Monday
Tuesday
Wednesday
Friday
qday

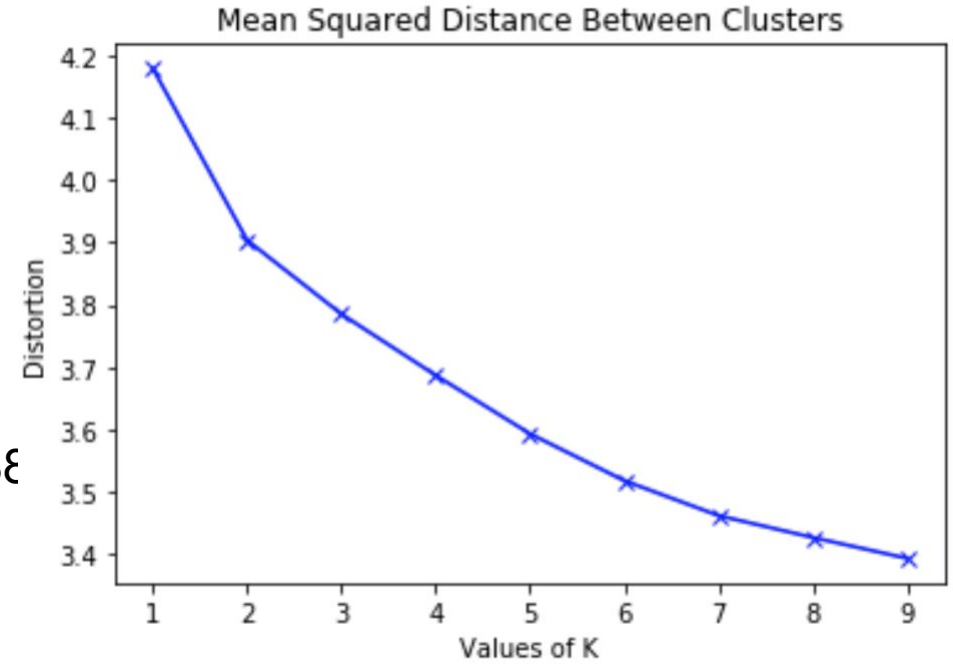
# CLUSTER

-Determine appropriate number of clusters
    - K = 2
-Cluster
    - HAC
    - **K-Means**

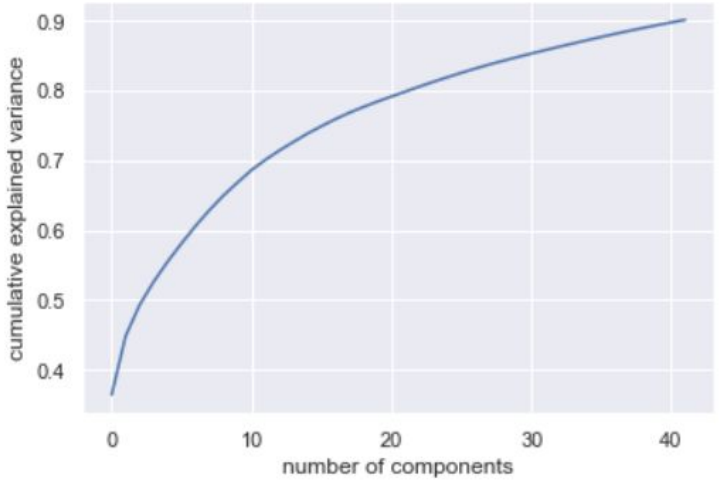
-Evaluate

    -Silhouette Score: 0.31

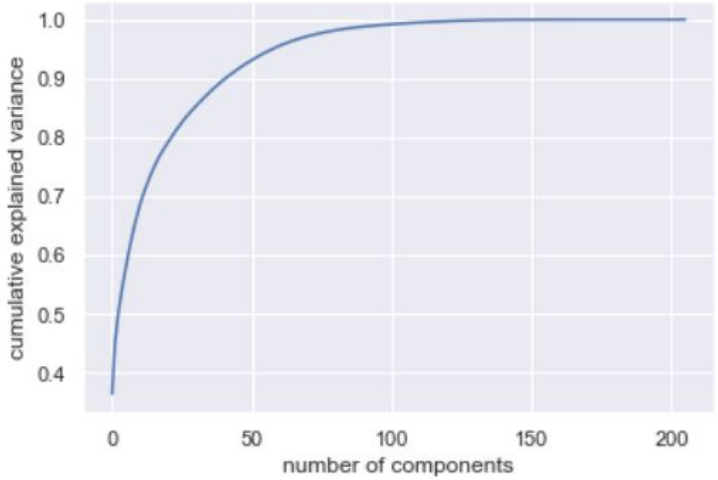
    -Calinski-Harabasz Score: 1491.88



# CLUSTER



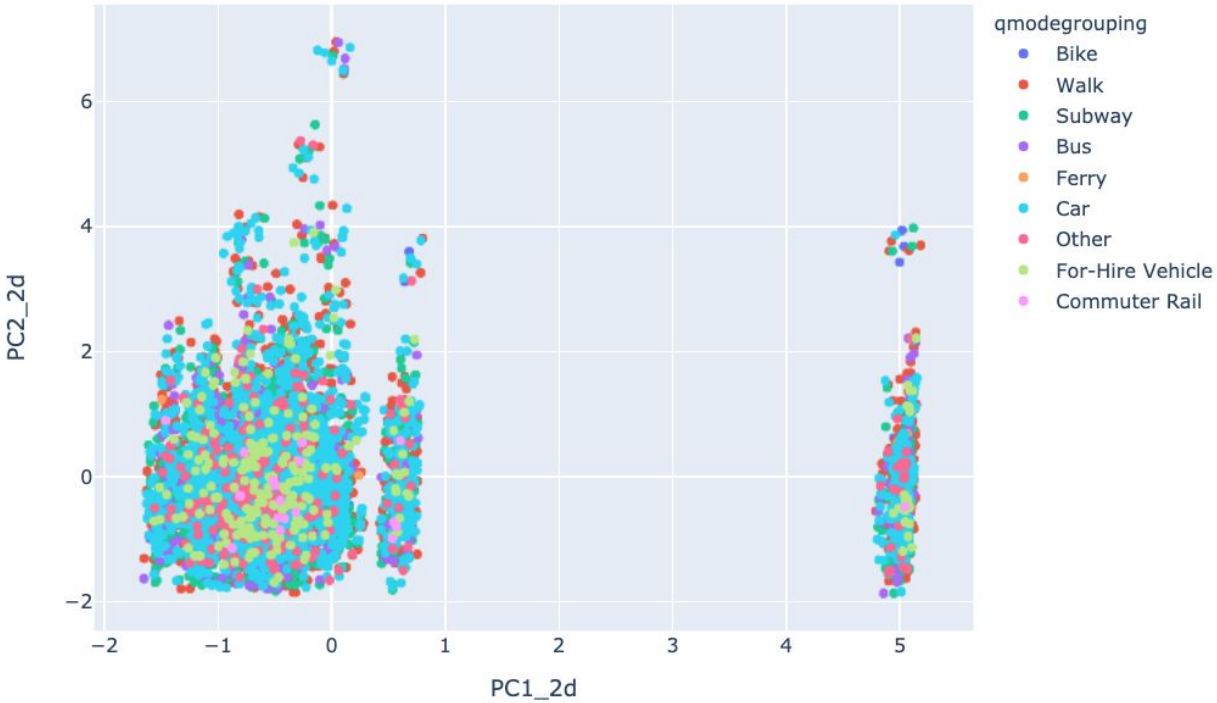

- 53% of variance can be explained by 4 principal components



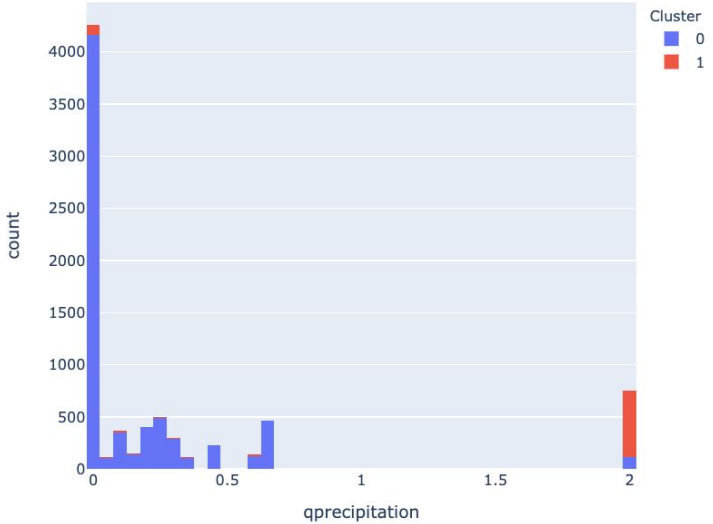

- 90% of variance can be explained by 42 features

# CLUSTER

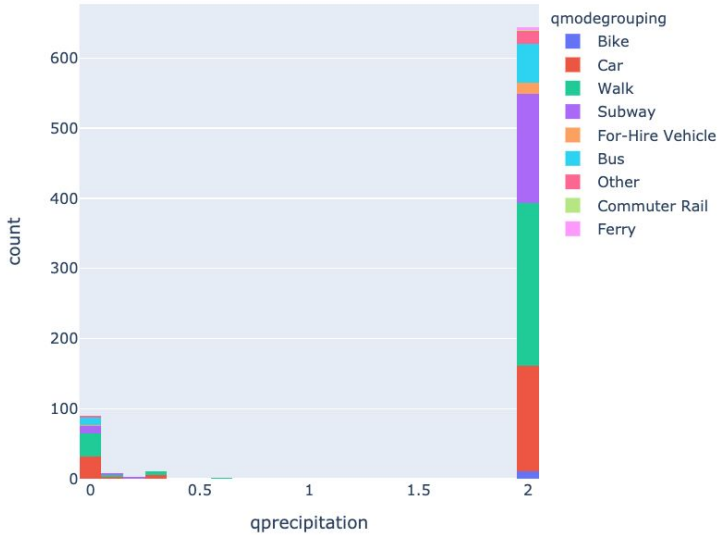
- K-Means where k=2
- Visualized with 2 Principal Components



# EDA: CLUSTER 1




- 85% of Cluster 1 trips occurred on days with 2in of precipitation




- No single form of transportation dominates in Cluster 1

# RECOMMEND!

- Build content-based recommendation engine for each cluster
- Offer alternative transportation for trips taken with cars

```
1  recommendations1(53202320)
```

```
[13203628,
 13203753,
 13203574,
```

```
1  min_cluster1_df.ix[(min_cluster1_df.trip_id =='13203628').idxmin(), 'qmodegrouping']
2
```

```
'Bike'
```

# THANK YOU!