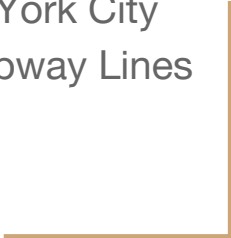# Yellow Taxi or Subway?

Big Data Analysis On New York City
Yellow Cab Trips Along Subway Lines

Background

vs.

In this project, we would find out the subway line with the most trips by taxi which could have been done by taking that line.
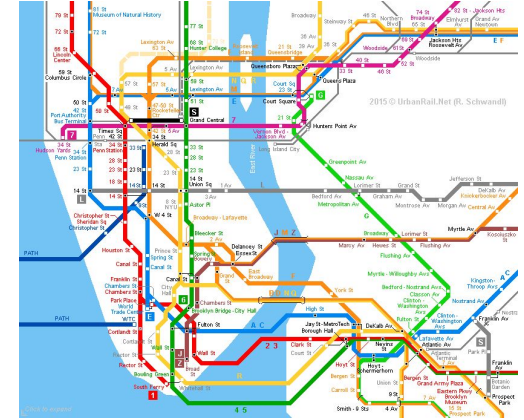
Questions

1. Any anomaly day or anomaly time?

2. Any seasonal pattern by line?

3. Which line got 'stolen' business most by yellow cab?

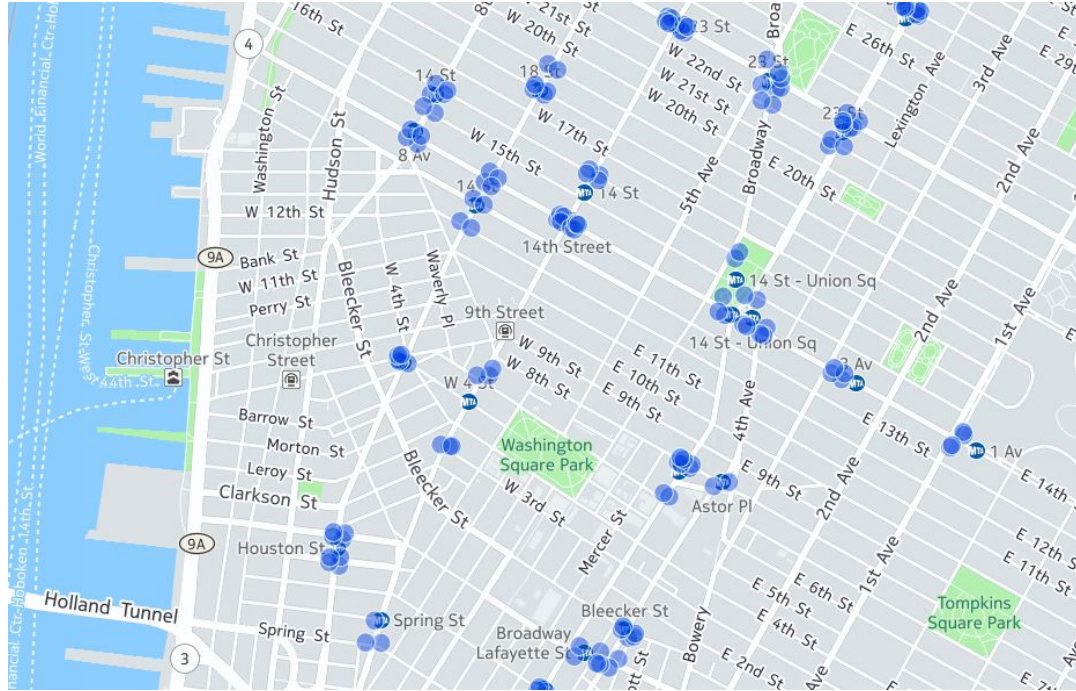4. Along which lines yellow cab trips show similar pattern?

Data

TLC Trip Record data
Yellow Taxi Cab
From July 2015 to June 2016
1.6 GB x 12 months

2016 (May) New York City Subway Station Entrances
From GIS Lab of CUNY
Presenting 1868 subway station entrances

2016 (May) New York City Subway Routes
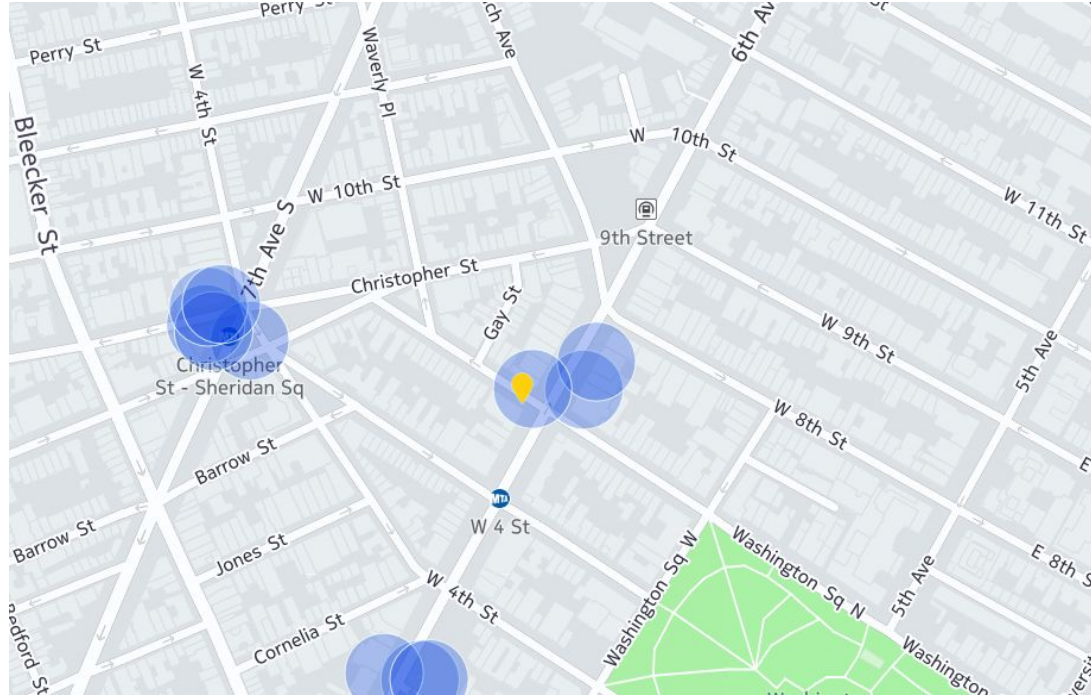From GIS Lab of CUNY
Showing 22 subway lines

Pre-Processing

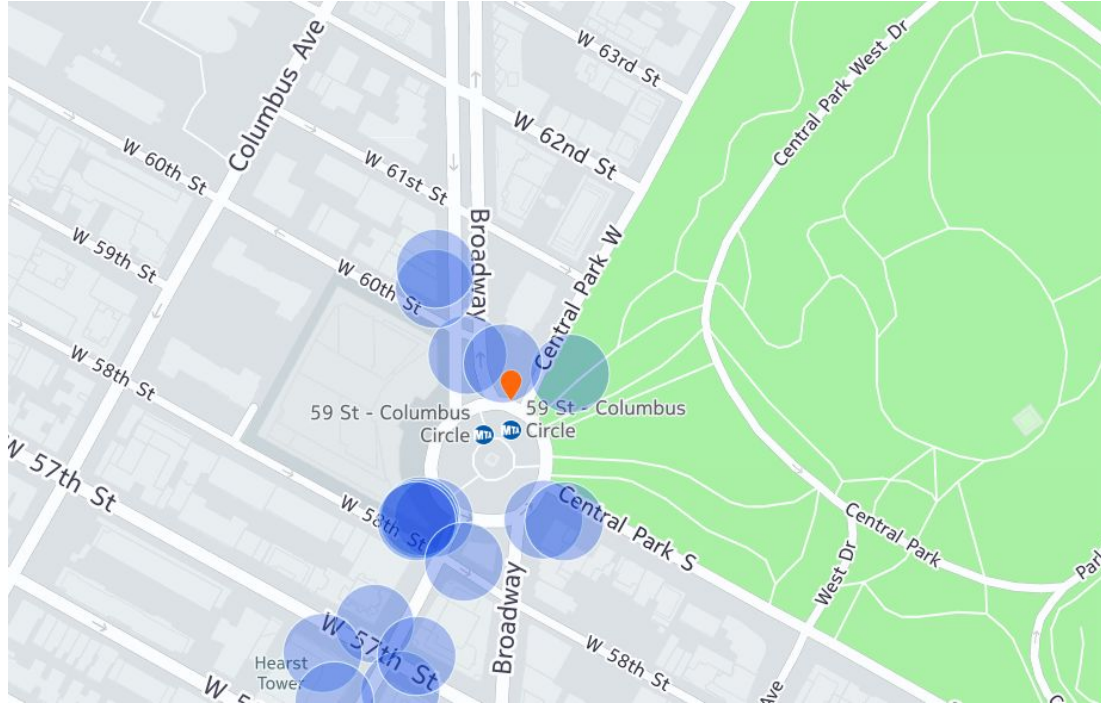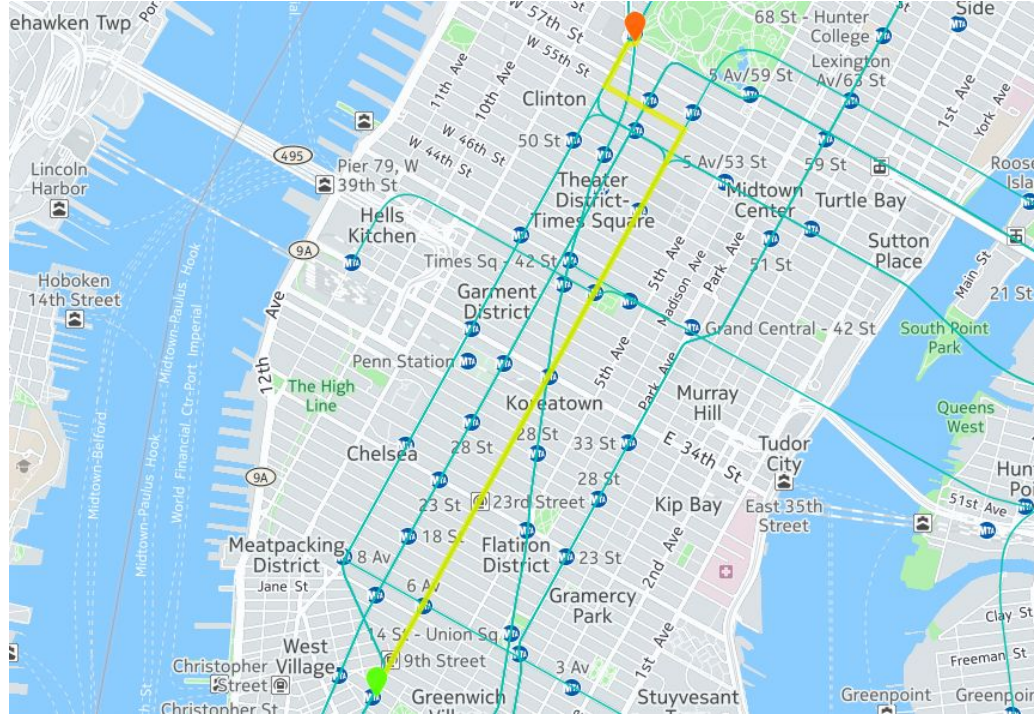Buffer 100 us-ft (30 m) of Subway Entrances

Example

Pre-Processing

West 4th St
NW Entrance

Pre-Processing

59 St-Columbus Circle West Entrance

Pre-Processing

Taxi Route

Subway Lines

Pre-Processing

Entrance	Line
West 4th St NW	A, B, C, D, E, F & M
59 St - Columbus Circle West	A, B, C, D & 1

Pre-Processing

This map depicts weekday service.
On weekends and late nights, these routes change:

Weekends

B	**M**	**N**	**W**	**5**
No service- use C D Q	Delancey St/ Essex St– Metropolitan Av	Local in Manhattan, via Manhattan Bridge	No service- use N R	Dyre Av– Bowling Green

Late nights (midnight to 6am, daily)

A	**C**	**E**	**B**	**D**	**M**	**S** 42 St
Local, 207 St– Far Rockaway. Euclid Av-Lefferts Blvd Shuttle	No service- use A	Local service	No service- use A D Q	Local in Brooklyn	Myrtle Av– Metropolitan Av Shuttle	No service- use 7

N	**Q**	**R**	**W**	**2**	**3**	**4**	**5**
Local, via Financial District	Local service	Whitehall St- 95 St	No service- use N	Local service	148 St– Times Sq/42 St	Local, Woodlawn– New Lots Av Skips Hoyt St	Dyre Av– E180 St Shuttle

Pre-Processing

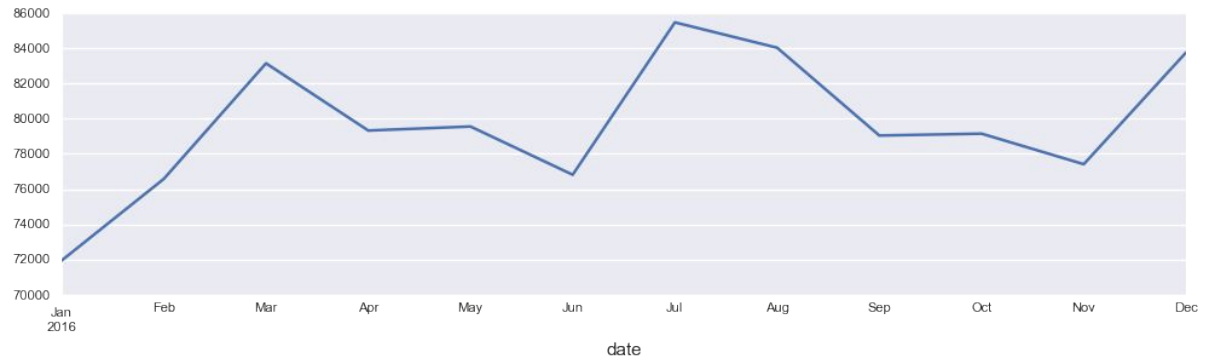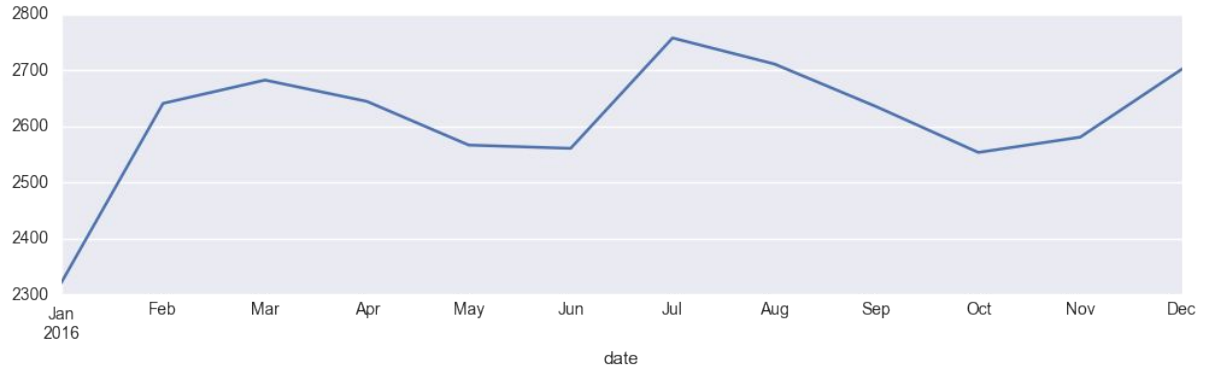| Line	Count
A	1
B	1
C	1
D	1

Output Analysis

- Monthly Pattern

- Seasonal Detection
  - General Pattern
  - Compare with mean

- Anomaly Detection
  - Anomaly Day
  - Anomaly Hour

- Line Pattern
  - Which line is most undesired
  - Which lines shared same pattern

Monthly Pattern

Mostly people choose to

take taxi other than subway in

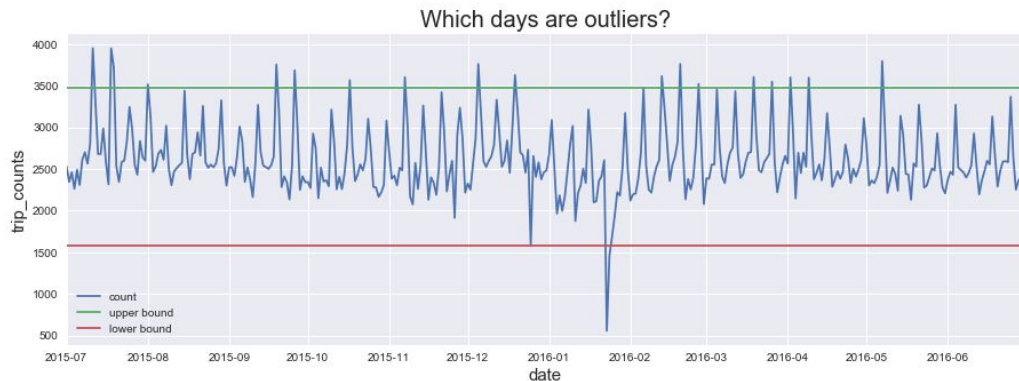July and December.

Sharp decrease in Jan. Why?
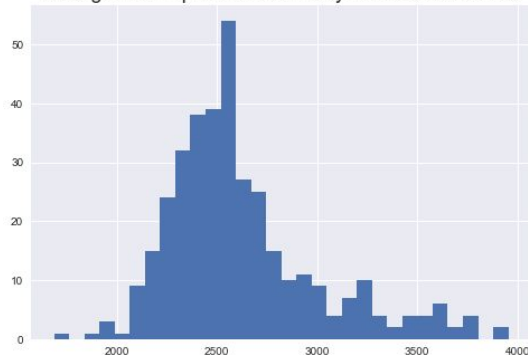
Seasonal Pattern by Line

- Appears that people will choose yellow taxi than subway in summer and winter.
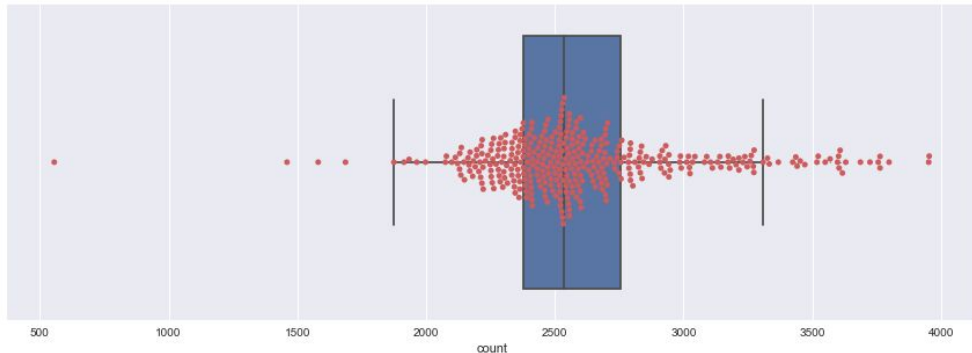- The C line has the most difference between seasons, especially summer and winter.

Anomaly Detection- Day

Which days are outliers?

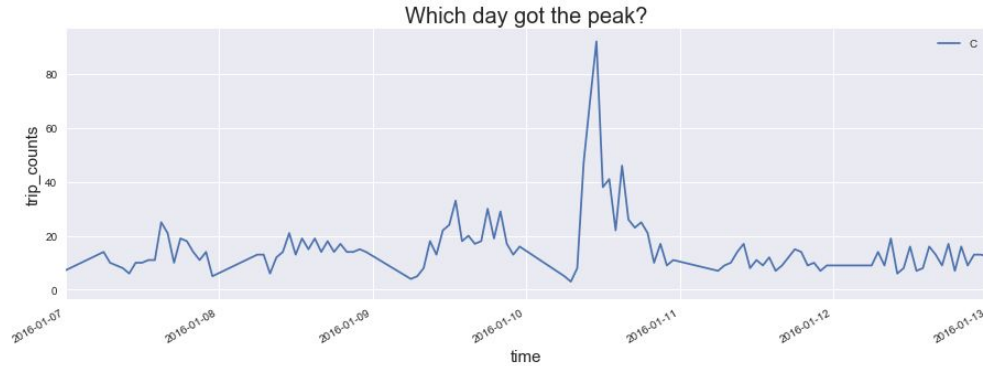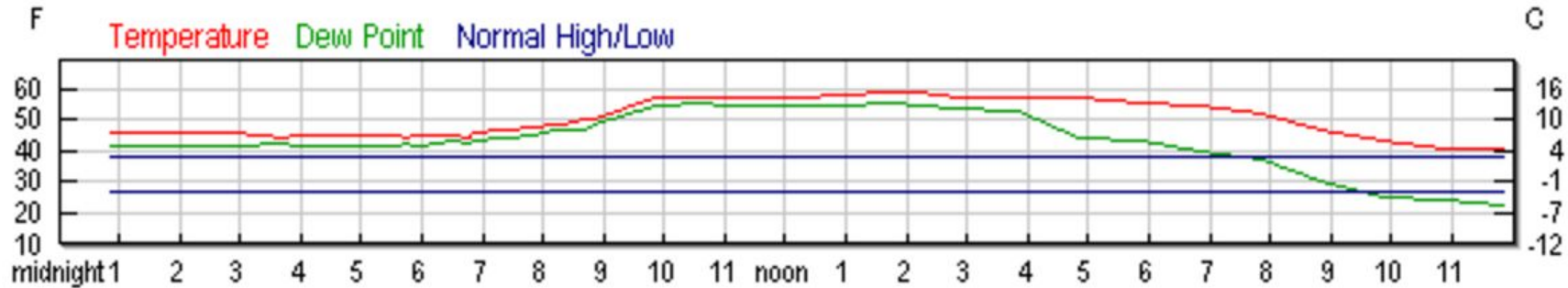Histogram of trip counts from July 2015 to June 2016

Which are outliers?

2016 Blizzard Was NYC's Biggest Snowstorm on Record, NOAA Report Finds

Snowfall totals in Central Park were upped from 26.8 inches to 27.5 inches, making the Jan. 22-23 storm the biggest blizzard to hit the city since recordkeeping began in 1869, according to the National Oceanic and Atmospheric Administration.
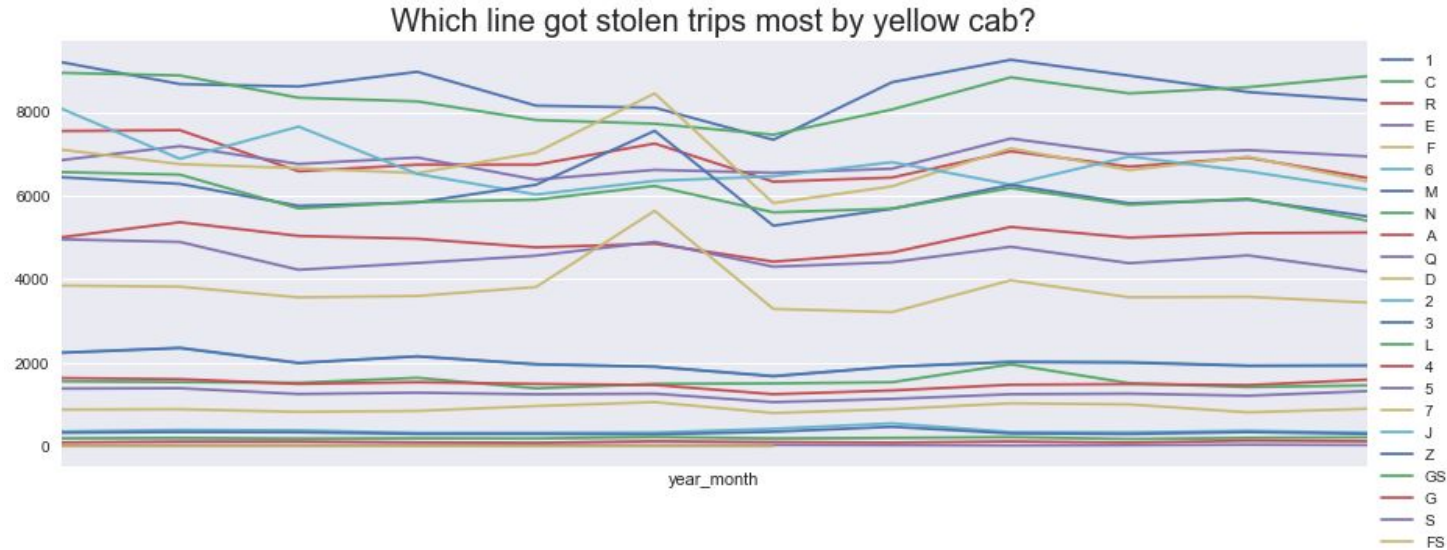
Anomaly Detection- Hour

Which day got the peak?

Which line got 'stolen' trips most by yellow cab?
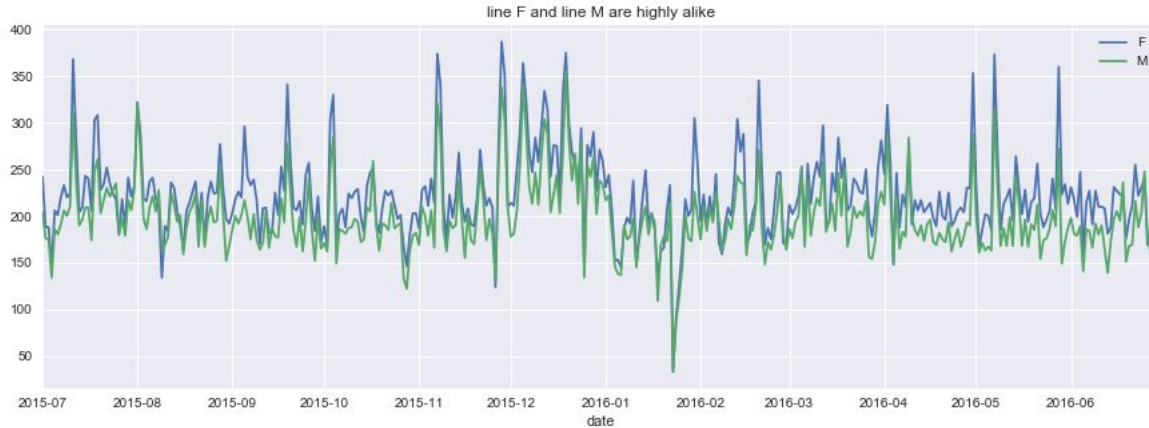

Which line got stolen trips most by yellow cab?

- Two Sample T-test shows, difference of line 1 & line C is not significant; difference OF line C & line R is significant.
- So line 1 and line C both got 'stolen' trips most by yellow cab.
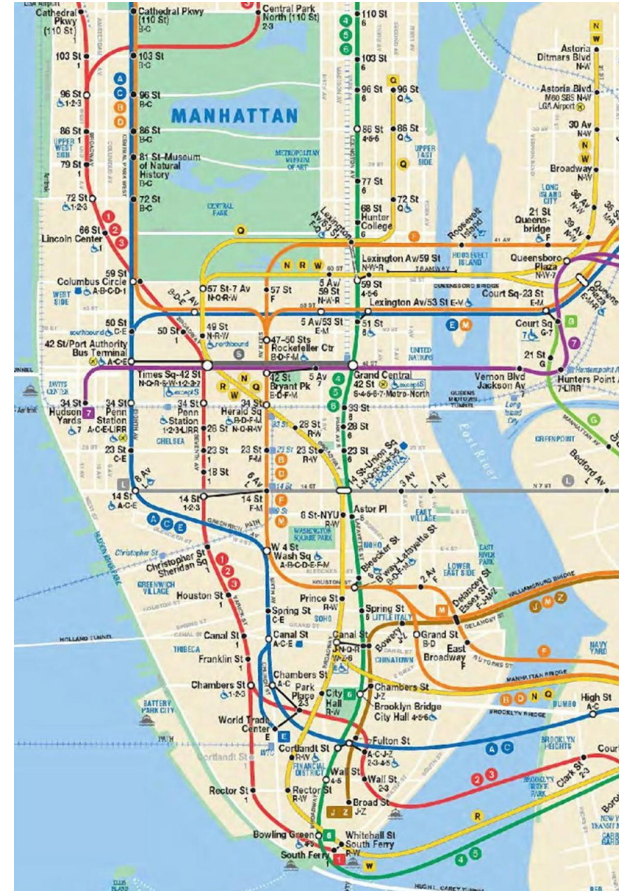
Which Two lines are similar?

Pearson correlation test shows:
(2,3), (J,Z), (N,R), (N,Q), (Q,R), (5,4), (F,M) has p-value less than 0.05, and coefficient more than 0.9.

line F and line M are highly alike

(F,M) shows Broadway/Lafayette st to 47-50st Rockefeller Center is significant for yellow cabs.

Conclusion

- People use yellow cab as a transport mode when go for recreation, not for working commute.
- When weather is pleasant, more people go out with yellow cab; when snow storm comes, the trips drop, maybe because drivers are not working as well.
- Saturday has significant peak on yellow cab trips.
- People will prefer yellow taxi in winter and summer compared to spring and fall.
- Broadway/Lafayette st to 47-50st Rockefeller Center is the active area people using yellow cab a lot, even if subway is really accessible as well, maybe due to tourists.
- Line 1 and C got 'stolen' business most by yellow cab.

Future

- Different population density at each subway station

Normalization to minimize the effect of different population density

- The length of subway lines