



DineSafe

SCS 3250-021 Foundation of Data Science

Raaid Batarfi, Jiarun He, Hwarang Kim, Mehdi Ostadhashem, David Scott



Agenda

Background

Data Preparation

Analysis

- Infraction Ratio
- Pass rate
- Establishment Safety Score

Challenges

Conclusion



Agenda

Background

Data Preparation

Analysis

- Infraction Ratio
- Pass rate
- Establishment Safety Score

Challenges

Conclusion



Background

- DineSafe: is a system focus on the inspection of all establishments serving and preparing food in the city of Toronto
 - Pass
 - Conditional pass
 - Closed notice
- Every establishment receives 1-3 inspections per year
- Three risk levels is used to categorize each establishment
 - High risk
 - Moderate risk
 - Low Risk



Agenda

Background

Data Preparation

Analysis

- Infraction Ratio
- Pass rate
- Establishment Safety Score

Challenges

Conclusion



Data Preparation

- Source of the data: The city of Toronto open data catalogue website
 - 90,520 records
 - 16,291 establishments
 - 25 months (from September 2016 to September 2018)
- The quality of the data was good
 - Most columns contained Non-numeric values which were treated as strings
 - There were several missing values or details
 - These were intentionally left when there was no infraction, no court order or no fine.
- We created extra columns from the inspection data column
 - Year , Quarter, Month, Week, Year & Quarter, Year & Month, Year & Week and Week & Day



Agenda

Background

Data Preparation

Analysis

- Infraction Ratio
- Pass rate
- Establishment Safety Score

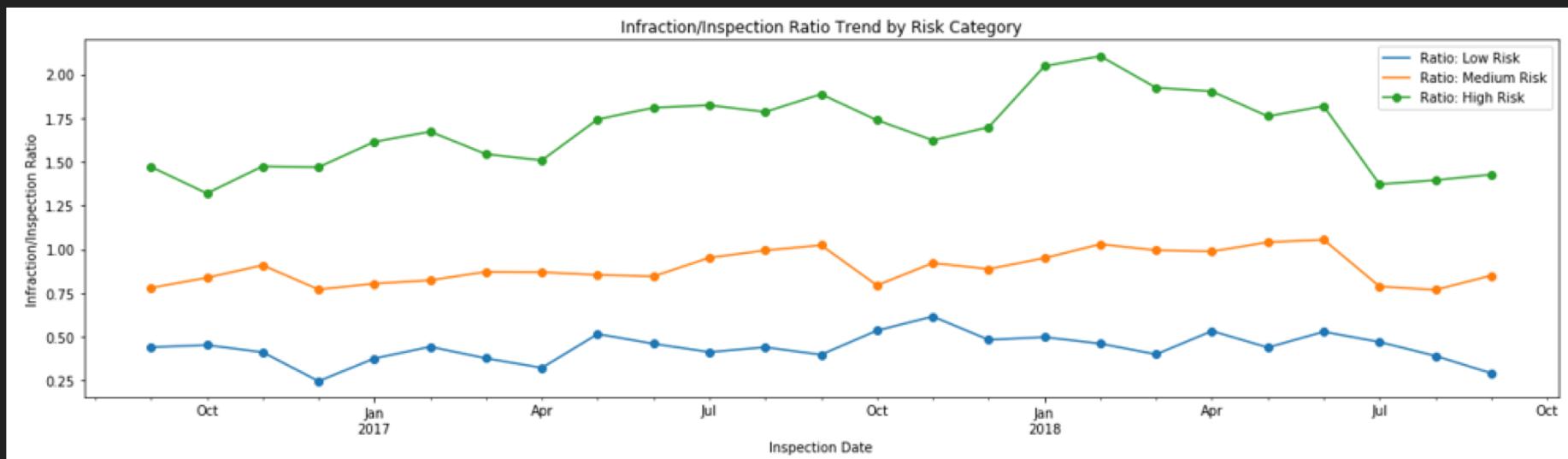
Challenges

Conclusion



Infraction Ratio

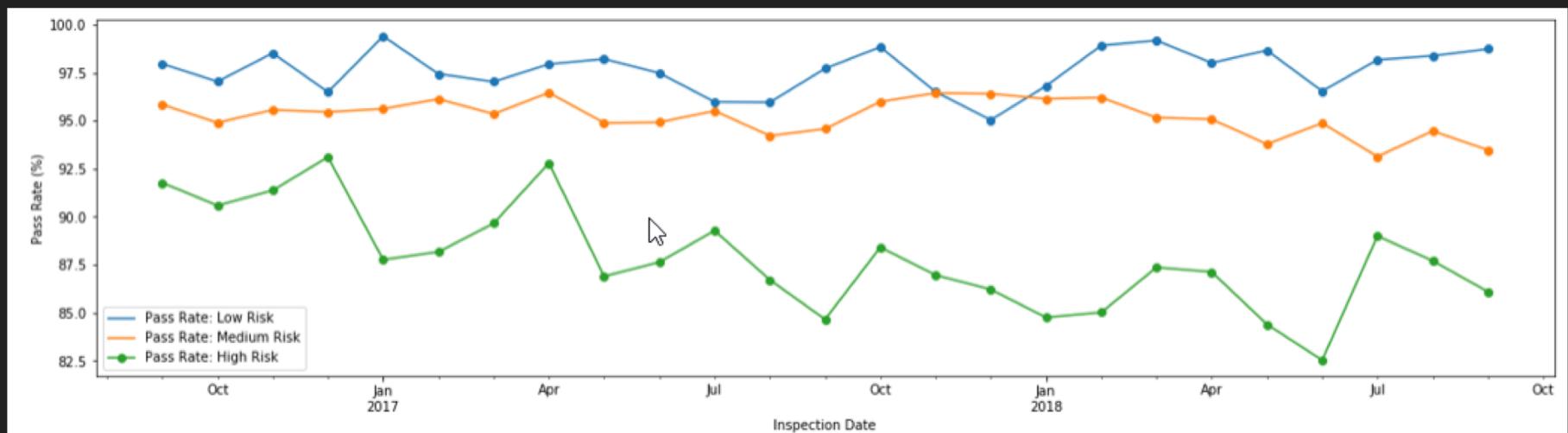
High risks establishments are more likely to have multiple infractions per inspection.





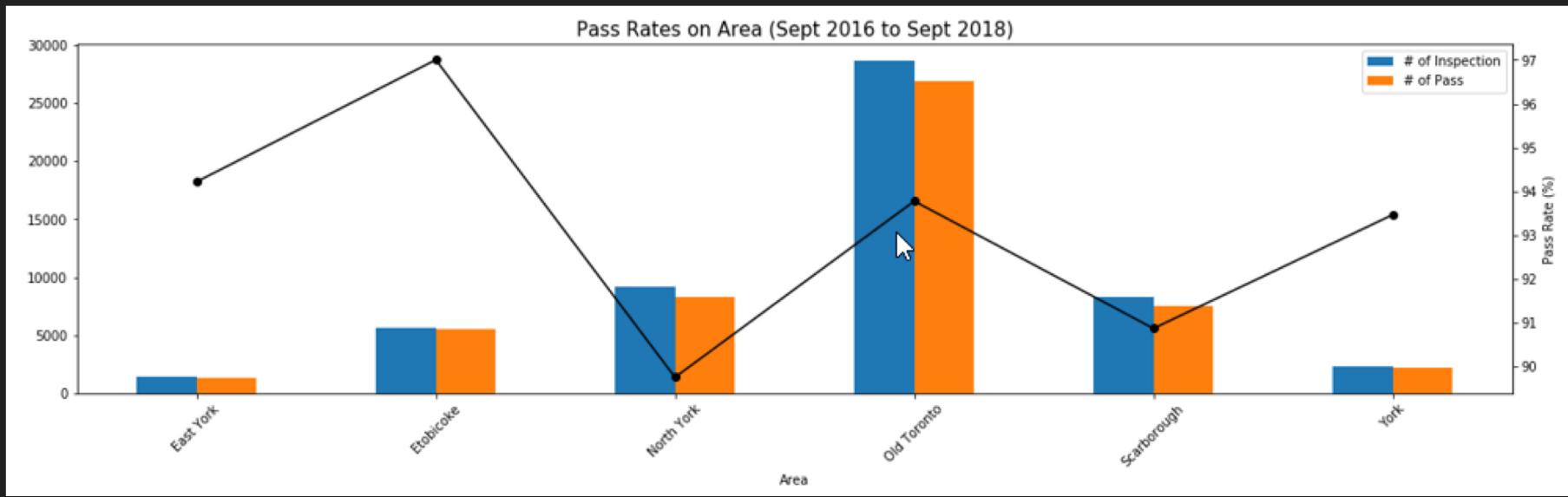
Pass Rate

The decreasing trend in pass rate can be attributed to a drop in the passing rate of high risk establishments



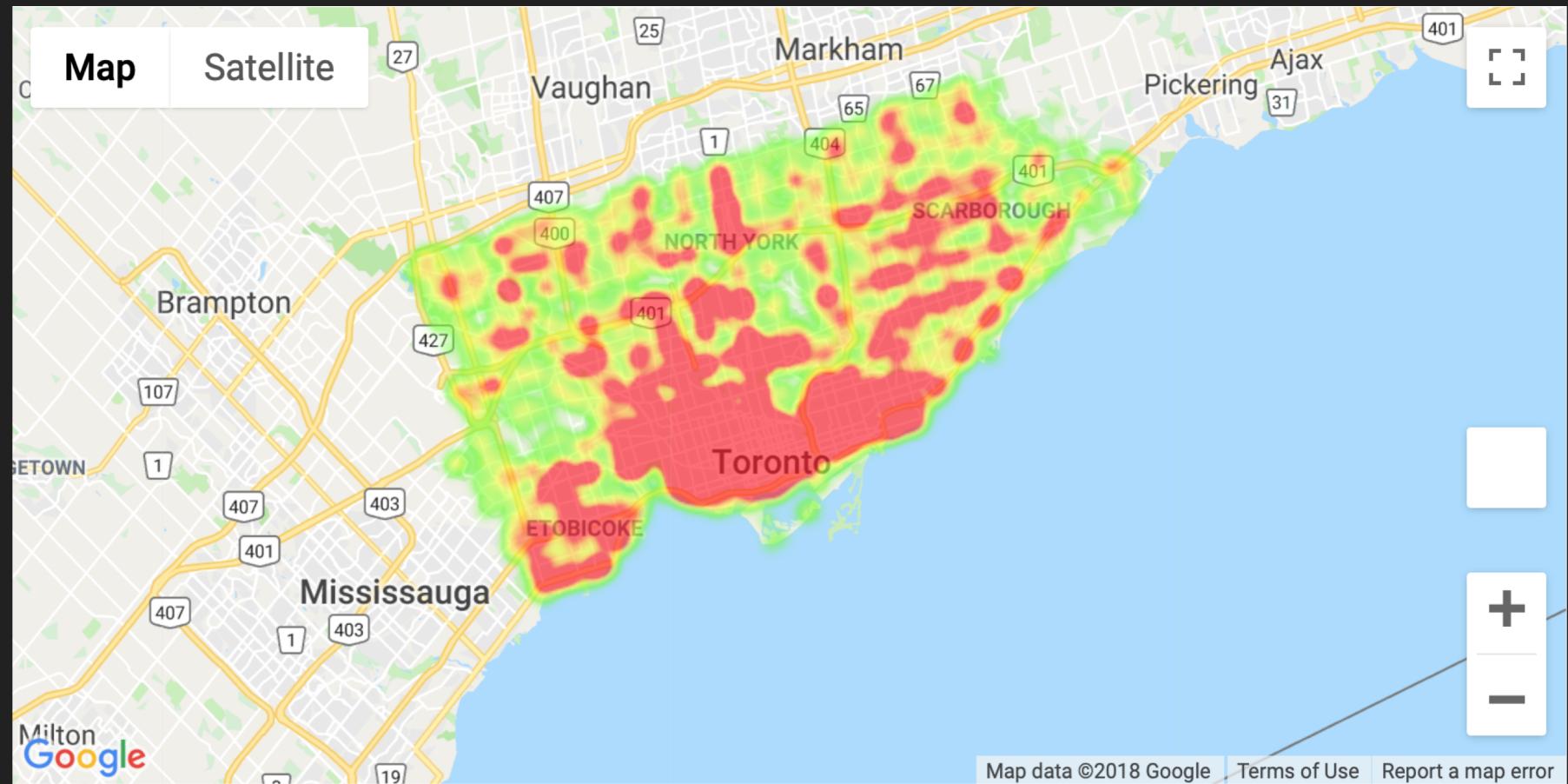
Pass Rate

What are the safest food preparation establishments? Are some areas safer than others?



Map

Satellite



Map data ©2018 Google

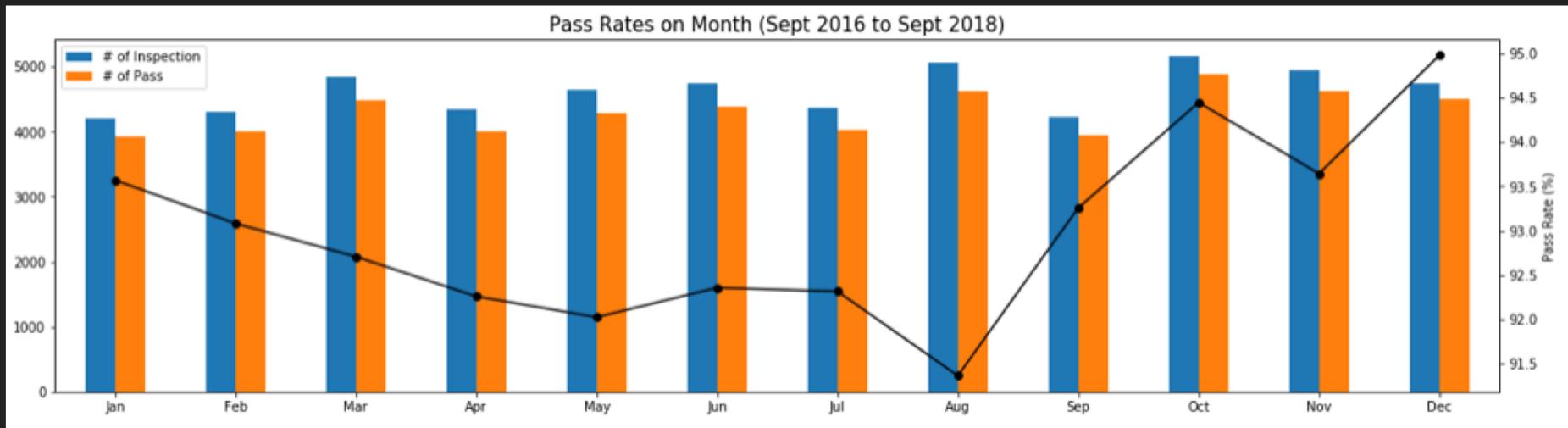
Terms of Use

Report a map error



Pass Rate

Seasonality and day of the week have an impact on the inspection result.



Establishment Safety Score

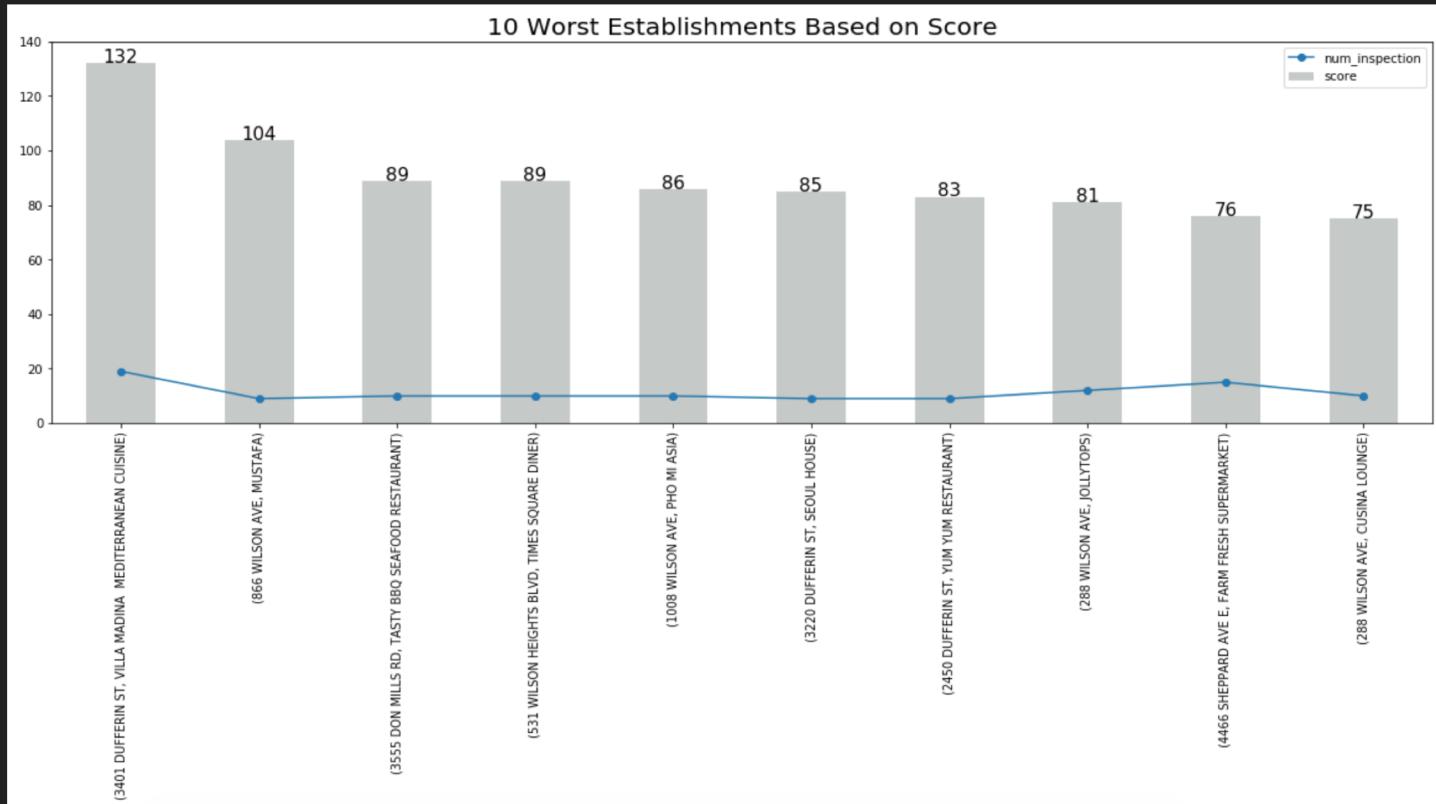
Three types of infractions:

- **Minor (score = 1)**
 - Minimal risk to public
 - Must be corrected in the next inspection
- **Significant (score = 2)**
 - Can lead to safety hazards
 - Must be corrected within 24-48 hours
- **Crucial (score = 3)**
 - Dangerous
 - Immediate health hazards
 - Must be corrected on the spot / “order to close” will be issued

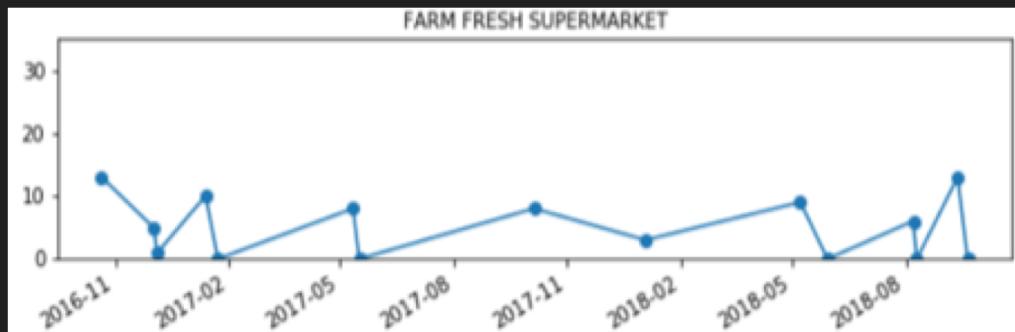
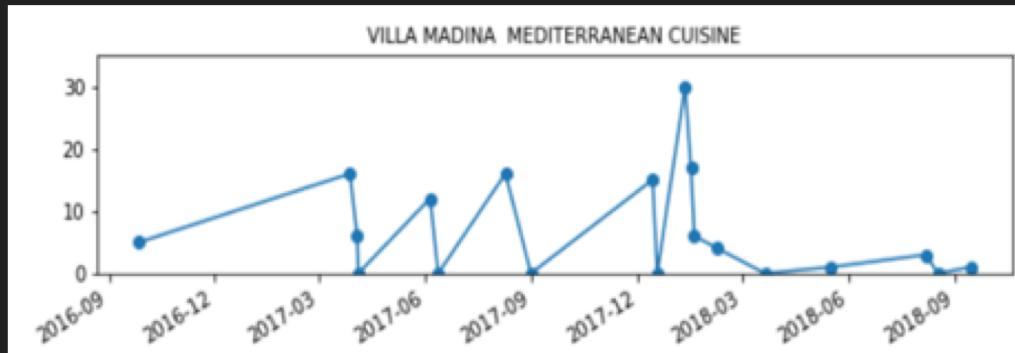
Establishment Safety Score Result

establishment_address	establishment_name	freq_per_year	num_inspection	score	Avg_Score
3401 DUFFERIN ST	VILLA MADINA MEDITERRANEAN CUISINE	9.5	19	132	6.9
866 WILSON AVE	MUSTAFA	4.5	9	104	11.6
3555 DON MILLS RD	TASTY BBQ SEAFOOD RESTAURANT	5.0	10	89	8.9
531 WILSON HEIGHTS BLVD	TIMES SQUARE DINER	5.0	10	89	8.9
1008 WILSON AVE	PHO MI ASIA	5.0	10	86	8.6
...
1027 STEELES AVE W	KIVA'S BAGEL BAKERY & RESTAURANT	4.0	8	68	8.5

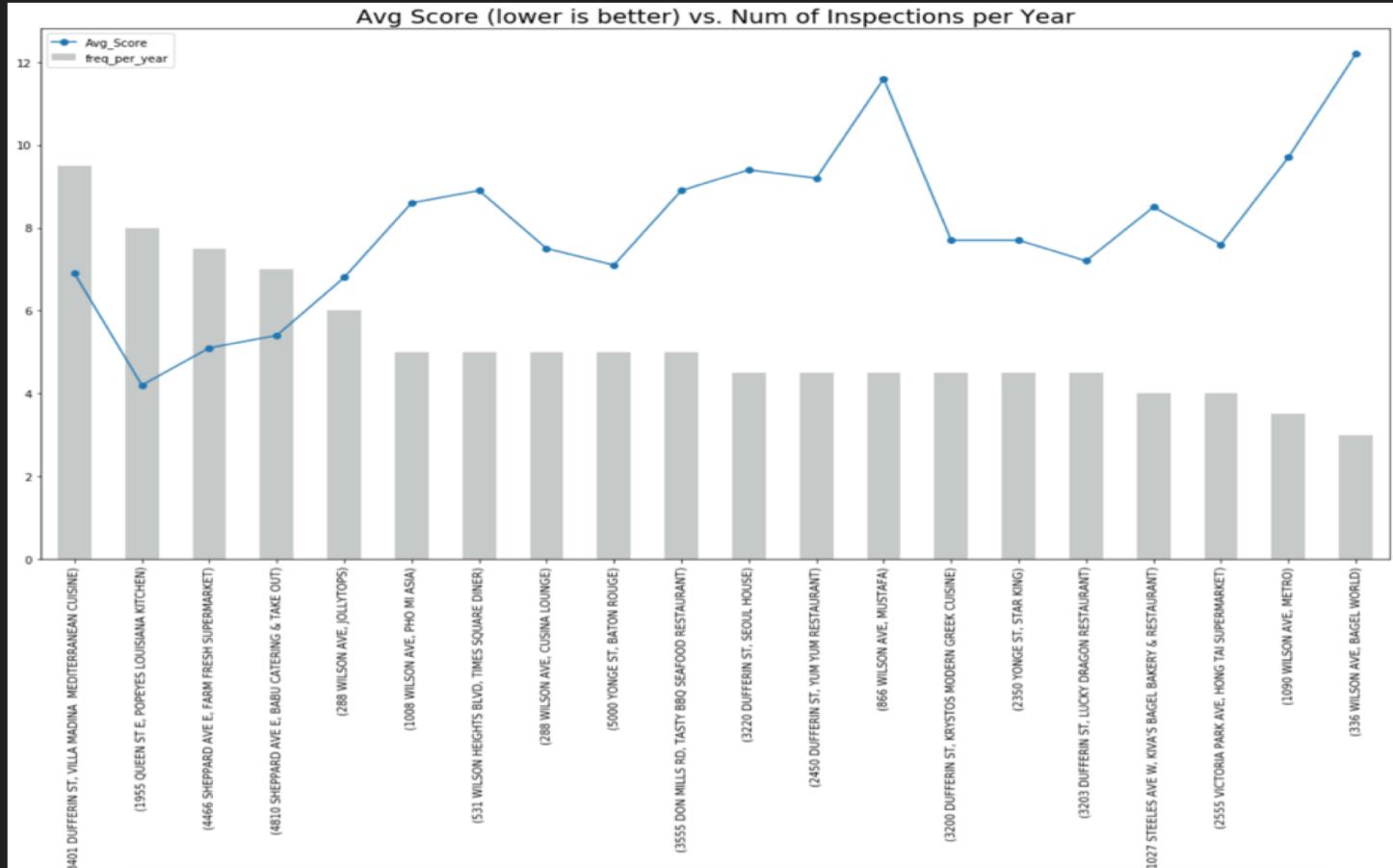
10 worst establishments based on score



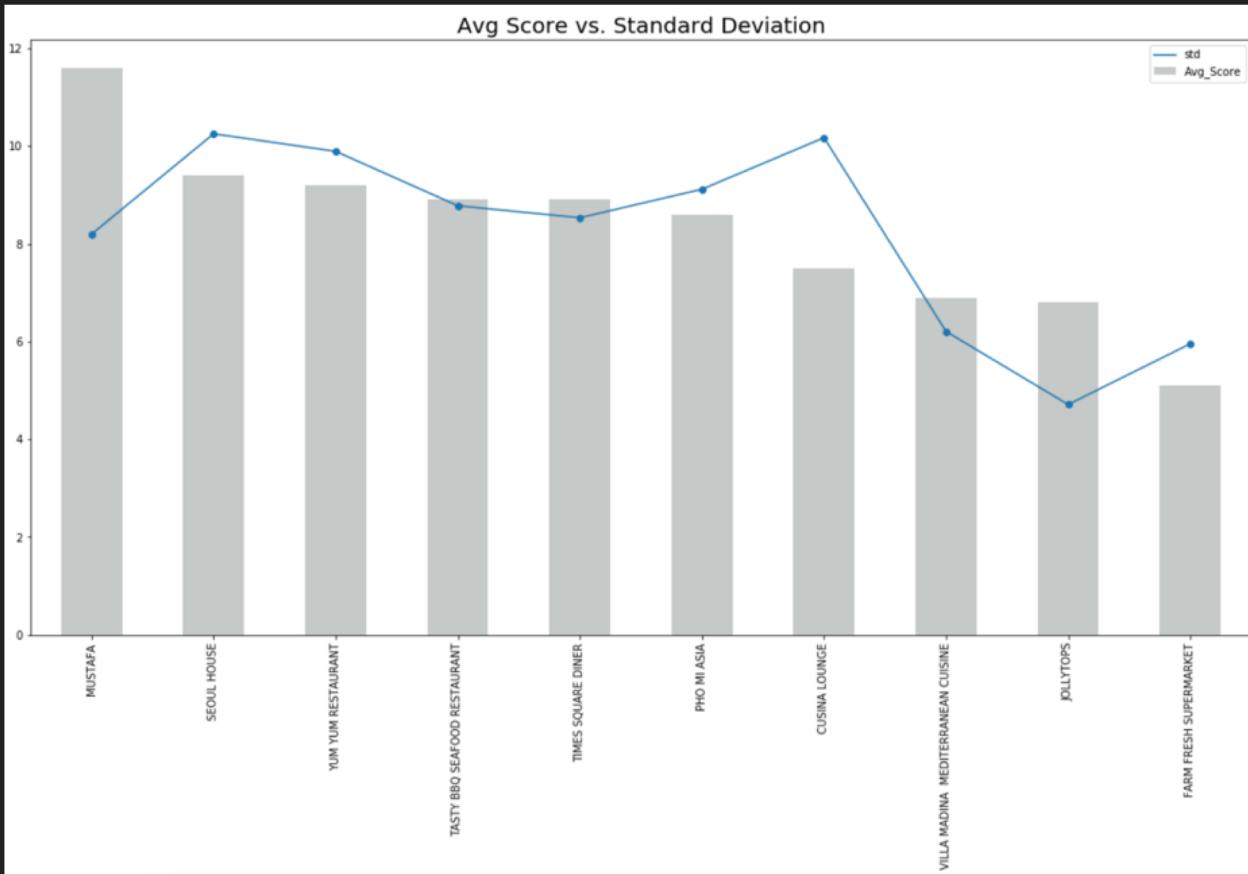
Villa Madina & Farm Fresh Time Series of Safety Score



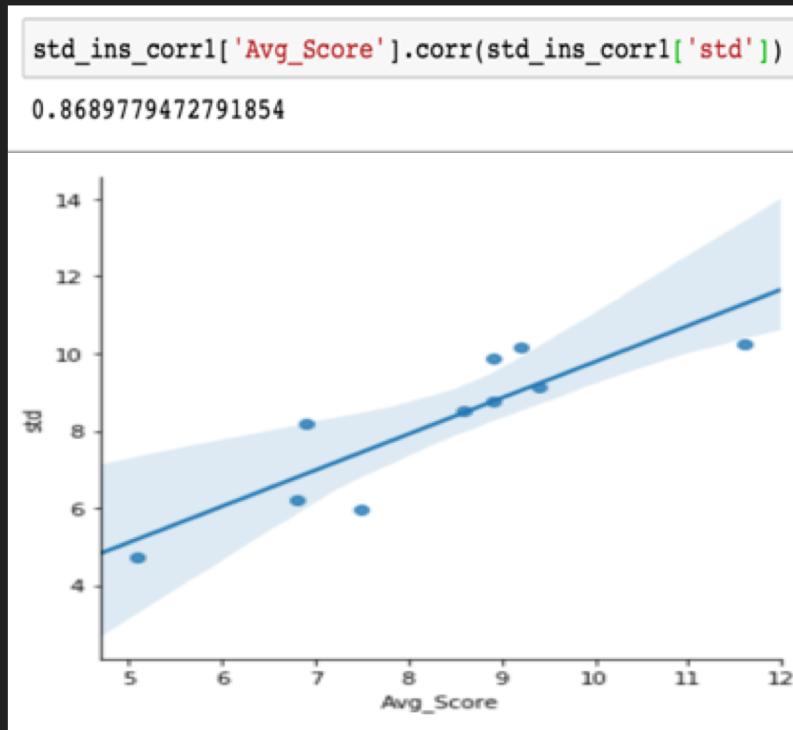
Number of inspections vs. safety score



Statistical Analysis of the Establishment Safety Score



Statistical Analysis of the Establishment Safety Score





Agenda

Background

Data Preparation

Analysis

- Infraction Ratio
- Pass rate
- Establishment Safety Score

Challenges

Conclusion



Challenges

XML file to CSV

Outliers in latitude and longitude columns

No area or postal codes (google API is used to convert)

- 100K records, 6K processing rate per hour
- Create unique address table, and separate data frame



Agenda

Background

Data Preparation

Analysis

- Infraction Ratio
- Pass rate
- Establishment Safety Score

Challenges

Conclusion



Conclusion

- Infraction Ratio
- Pass rate
- Establishment Safety Score
- More inspections → lower infraction score
- More consistent food safety control → lower infraction score