Analytics and Observed Trends of Issued Parking Tickets in the New York City Metropolitan Area

Hannah VanderHoeven, Ryan Smith, Angel Augustine

CS 455: Introduction to Distributed Systems

Computer Science Department, Colorado State University

Background Information

- New York City has the highest population and population density in the country, meaning there is competition for parking.
- In Manhattan alone, the population as of 2019 was 1.629 million with a land area of 22.822 miles (e). This translates to approximately 71,000 people per square mile!
- With so many people in such a highly concentrated area, there’s limited space for parking; specifically, street parking.
- Also different violations vary in price so understanding parking ticket trends could be very useful to someone trying to save money.

Problem Characterization

- By analyzing millions of parking ticket entries from the year 2016, conclusions can be drawn to help drivers discover scenarios where they might be at a higher risk for getting a parking ticket.
- Specifically looking at locations where most tickets are given out, the types of violations that are given the most in these locations, the time of day most tickets are given out, and characteristics of vehicles that were given the most tickets.

Methodology - Part 1

- Spark job was used to parse and analyze the average likelihood of multiple different scenarios, along with the total values over the year for other values that could be used to around out the dataset.
- This was done in multiple parts, first by parsing the data in the CSV file out into useable Scala data models, also known as case classes.
- The TicketData model was made up of everything in the CSV file, this included all of the 51 different fields that mapped everything from the date the ticket was issued, to the vehicle information and the ticket issuer.

```scala
//The ticket data model
case class TicketData(SummonsNumber: String, PlateID: String, RegistrationState:String, PlateType: String,
                      Date: String, ViolationCode: String, VehicleBodyType: String,
                      VehicleMake: String, IssuingAgency: String, StreetCode1: String,
                      StreetCode2: String, StreetCode3: String, VehicleExpirationDate: String,
                      ViolationLocation: String, ViolationPrecinct: String, IssuerPrecinct: String,
                      IssuerCode: String, IssuerCommand: String, IssuerSquad: String, ViolationTime: String,
                      TimeFirstObserved: String, ViolationCounty: String, ViolationInFrontOf: String,
                      HouseNumber: String, StreetName: String, IntersectingStreet: String,
                      DateFirstObserved: String, LawSection: String, SubDivision: String,
                      ViolationLegalCode: String, DaysParkingInEffect: String, FromHoursInEffect: String,
                      ToHoursInEffect: String, VehicleColor: String, UnregisteredVehicle: String,
                      VehicleYear: String, MeterNumber: String, FeetFromCurb: String, ViolationPostCode: String,
                      ViolationDescription: String, NoStandingStoppingViolation: String, HydrantViolation: String,
                      DoubleParkingViolation: String, Latitude: String, Longitude: String, CommunityBoard: String,
                      CommunityCouncil: String, CensusTract: String, BIN: String, BBL: String, NTA: String)
```

Parking_Violations_Issued_-_Fiscal_Year_2016.csv — Edited

```csv
Summons Number,Plate ID,Registration State,Plate Type,Issue Date,Violation Code,Vehicle Body Type,Vehicle Make,Issuing Agency,Street
Code1,Street Code2,Street Code3,Vehicle Expiration Date,Violation Location,Violation Precinct,Issuer Precinct,Issuer Code,Issuer
Command,Issuer Squad,Violation Time,Time First Observed,Violation County,Violation In Front Of Or Opposite,House Number,Street
Name,Intersecting Street,Date First Observed,Law Section,Sub Division,Violation Legal Code,Days Parking In Effect    ,From Hours In
Effect,To Hours In Effect,Vehicle Color,Unregistered Vehicle?,Vehicle Year,Meter Number,Feet From Curb,Violation Post Code,Violation
Description,No Standing or Stopping Violation,Hydrant Violation,Double Parking Violation,Latitude,Longitude,Community Board,Community
Council ,Census Tract,BIN,BBL,NTA
1363745270,GGY6450,99,PAS,07/09/2015,46,SDN,HONDA,P,0,40404,40404,20170602,0074,74,301,358160,T301,0000,1037A,,K,F,142,MACDOUNGH
ST,,0,408,D1,,BBBBBBB,ALL,ALL,WHITE,0,2010,-,0,,,,,,,,,,,,
1363745293,KXD355,SC,PAS,07/09/2015,21,SUBN,CHEVR,P,55730,67030,58730,20160288,0079,79,301,358160,T301,0000,1206P,,K,F,331,LEXINGTON
AVE,,0,408,F1,,YBBYBBB,1100A,1230P,RED,0,0,-,0,,,,,,,,,,,,
1363745438,JCK7576,PA,PAS,07/09/2015,21,SDN,ME/BE,P,42730,26730,26830,0,0079,79,0,358114,TEBN,0000,0820A,,K,F,1087,FULTON
ST,,0,408,D1,,YBBYBBB,0800A,0930A,WHITE,0,0,-,0,,,,,,,,,,,,
1363745475,GYK7658,NY,OMS,07/09/2015,21,SUBN,NISSA,P,58130,18630,67030,0,0079,79,301,358114,T301,0000,0918A,,K,F,207,MADISON
ST,,0,408,D1,,YBBYBBB,0900A,1030,BK,0,2015,-,0,,,,,,,,,,,,
1363745487,GMT8141,NY,PAS,07/09/2015,21,P-U,LINCO,P,58130,67030,58730,20160206,0079,79,301,358114,T301,0000,0925A,,K,F,237,MADISON
ST,,0,408,D1,,YBBYBBB,0900A,1030A,BLK,0,2006,-,0,,,,,,,,,,,,
1363745517,GYK3760,NY,PAS,07/09/2015,21,SUBN,HONDA,P,46730,58730,85730,20160709,0079,79,301,358114,T301,0000,0948A,,K,F,201,HALSEY
ST,,0,408,D1,,YBBYBBB,0900A,1030A,OTHER,0,2006,-,0,,,,,,,,,,,,
```

Methodology – Part 2

- The TicketLocation class was created as a subclass of the TicketData information, it contained Violation Location and Street Name.
- This combination of values was used as key in multiple of other data models to map more specific information for each of the top 10 ticketed locations.
- For each of these locations inner joins were used to create datasets of other characteristics that occurred frequently in the locations.
- These included average tickets given each day for certain vehicle colors, types, makes and violation types.

Methodology – Part 3

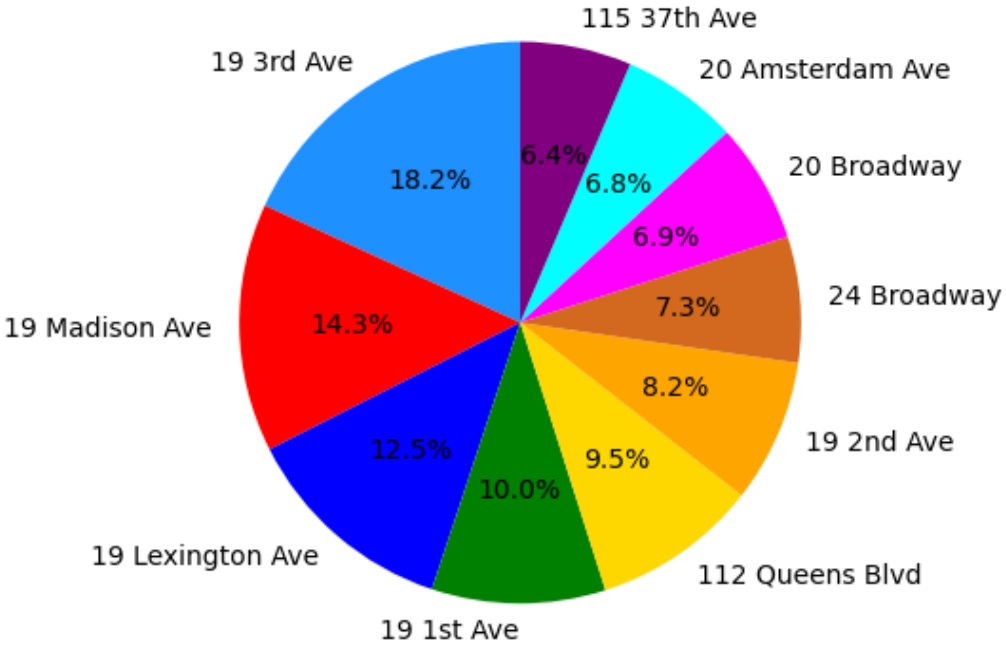
- Not all the data collected were averages, we also looked at some total values that spanned the entire year and had nothing to do with location.
- These included the total tickets given out by each precinct, total tickets given out at each time of day, and the registration state of each of the tickets given over the span of a year.
- These were simply found by parsing the string value for each out of the complete ticket list and reducing on this value to get the total count

Performance and Benchmarks – Part 1

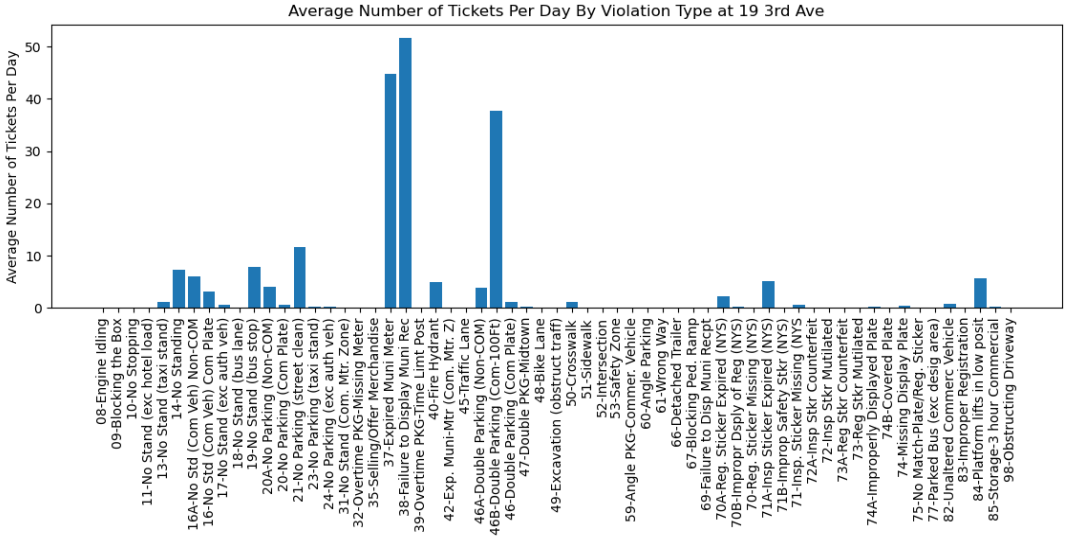
- Accuracy
  - When the tickets were broken down into the location objects, tickets where violation location or street name returned an empty value were then removed from the data set.
- Turnaround Time
  - Originally, the job took over a day to run and process the entire 2.15 GB dataset.
  - Using the “Persist” method to cache the data sets that were processed multiple times got the job running as quickly as two minutes and thirty seconds.
- Machines
  - The spark job was run on a cluster of 15 machines

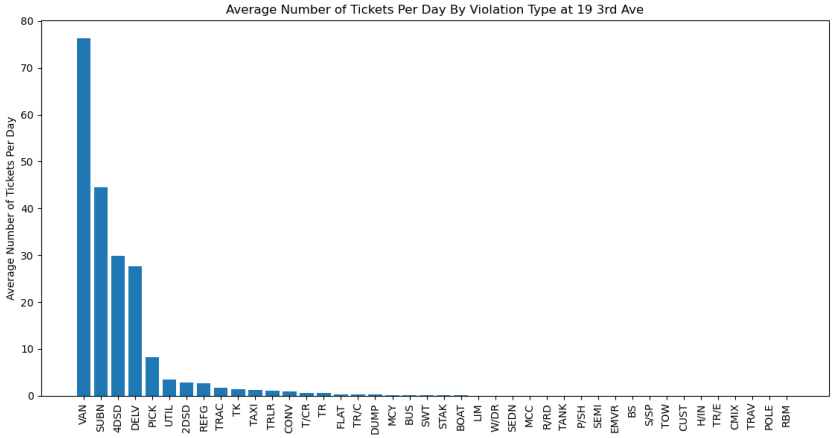
Performance and Benchmarks – Part 2

Top 10 Ticketed Locations, Daily Average

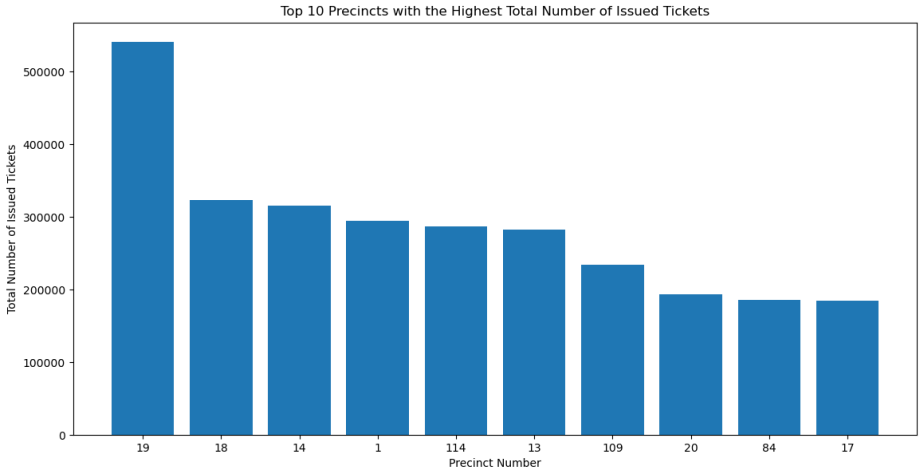


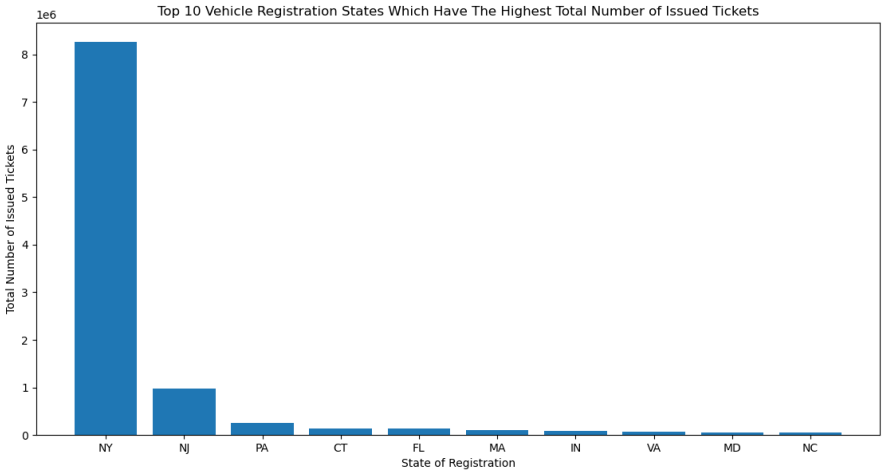
Daily Averages for the top ticketed location – 19, 3rd Ave.

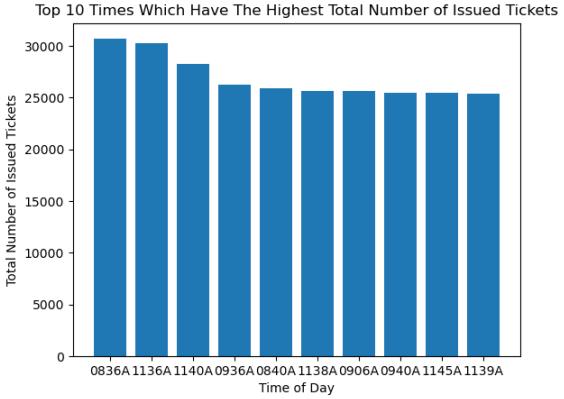




Performance and Benchmarks – Part 3







Insights and Conclusion

- Most tickets were given out in larger business areas, which makes sense since many people could get tickets while commuting from work or going to the businesses in the area.
- On average most tickets given out in the most ticketed areas were given to Vans, which could be explained by Vans being used the most by businesses in the area.
- Double parking and expired meter violations, occurred in higher frequencies in the top ticketed locations.
- Overall, knowing the trends of tickets issued is incredibly useful for drivers, but relying solely on one characteristic is not as helpful, as discovered in this project combining a variety of factors gives much more a of well-rounded picture.