911 Calls for Service Forecasting

Breonna Moore

Project Design

I used data from Open Baltimore to predict and forecast the number of calls received by Baltimore, MD Police Department. Insights gained from this model can be used by the police districts to plan their staffing schedule.

Forecasting for this project was performed using SARIMA as the naïve model and FB Prophet (Prophet). Root mean squared error was the metric chosen to compare the models because it produces an error that has the same units as the response variable. The SARIMA model is able to detect the weekly rise and fall trend, but it does not detect the seasonal trend of increased call counts during the summer. Prophet is able to detect both the weekly and the seasonal trend, and because of that, it produced the best model when optimizing for root mean squared error (RMSE).

Prophet's RMSE is for 2018 is 216 calls/day and SARIMA's RMSE for 2018 is 268 calls/day. By calculating the amount of money that can be saved using the using the average salary of $33/hour and the average time to service the calls of 18 mins, we find that Prophet can save $1.7M in one year in comparison to using the SARIMA model.

Tools

- Python
    - Data Collection: Pandas
    - Data Analysis: Scikit-learn, NumPy, StatsModels, FB Prophet
    - Data Visualization: Matplotlib, Seaborn, Tableau
- Microsoft PowerPoint, Microsoft Word

Data

There are 7 million calls for service in the dataset provided through Open Baltimore. The attributes of the data can be found in Appendix 1. In order to model the calls, they had to be resampled to list the entries and the counts by day. I also grouped the data by District because different districts answer calls for different areas. There was a strong monthly trend, but because I wanted to be able to predict the daily counts to allow district to guide their staffing using these numbers, I chose to forecast the daily counts. The call data consist of information about the calls, like time, date, and police district and information about why the call was placed like the kind of incident and its priority.

Tableau

The tableau dashboard uses python in the background and sends requests to python for the forecasted value based on the date and district. It the displays that forecasted value in a user-friendly design that allows the user to input a specific date and district.

What I Would Improve Upon Next Time

The difficulties with this project were that the 2019 year did not follow the same trends as previous years. I would have liked to explore using exogenous variables that would give some greater level of predictability around natural disasters. I also would have liked to choose my topic sooner, which would have allowed me more time to build out my tableau dashboard.

Appendix 1: Data

Variable	Description	Data Type
RecordID	Uniquely identifies a record or a call event	Number
CallNumber	Call reference number	String
CallDateTime	Date and time call was made	Date & Time
Priority	Priority of the call	String
District	Police district abbreviated	String
Description	Description of call	String
IncidentLocation	Street address of incident associated to call	String
ZipCode	Zip code of incident location	String
Neighborhood	Neighborhood name of incident location	String
PoliceDistrict	Police district name of incident location	String

PolicePost	Police post zone, call is falls in	String
CouncilDistrict	Council district call incidents falls in	String
SheriffDistricts	Sheriff districts call incident falls in	String
Community_Statistical_Areas	Community Statistical Areas call incident falls in	String
Census_Tracts	Census Tracts call incident falls in	String
VRIZones	Violent reduction initiative zones	String
Location	Geocoded location of the incident address	Location