Predicting Road Accidents in the City of Austin

Ramesh Adhikari

12/27/2019

Road Accident density in Austin:
Time period (Sep 26, 2017 to Sep 9, 2019)

Data Story

- Road accidents, a big problem
  68420 accidents in Austin, Texas in two years (September 2017 to September 2019)

- Machine learning can help to save by predicting the potential road accidents (time and location)

- We predict the number of daily and hourly accidents at different locations in the city of Austin

Data:

- Accidents count of accidents were collected every 5 second traffic incidence report from https://data.austintexas.gov/Transportation-and-Mobility/Real-Time-Traffic-Incident-Reports/dx9v-zd7x

- Hourly weather report was extracted from https://api.darksky.net.

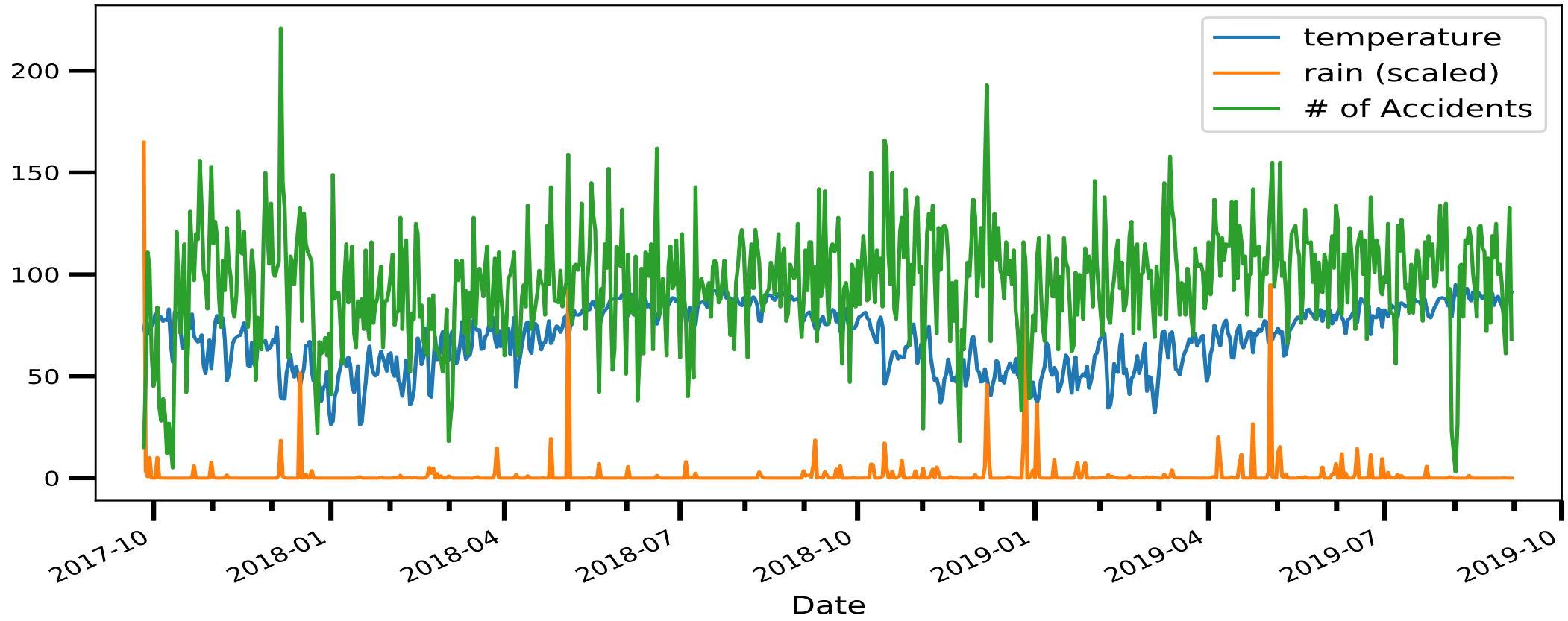
- Datasets related to schools, traffic signals, historical landmarks are collected from https://data.austin.gov,

Data Features:

Column1	Time	year	month	day	hour	DayOfWeek	Holiday	Latitude	Longitude	Location	Weather	pcpt_mmph	visibility	humidity	windSpeed	Traffic_signal	Schools	Landmarks	Solar_inclination	temperature	dewPoint	Issue Reported	Accident
0	11/17/17 14:00	2017	11	17	14	4	0	30.214	-97.83032		partly-cloudy-day	0	9.997	0.61	6.35	64	31	2	34.69635697	81.96	67.03	Crash Urgent	1
1	11/17/17 14:00	2017	11	17	14	4	0	30.21379	-97.83035		partly-cloudy-day	0	9.997	0.61	6.35	64	31	2	34.69655145	81.96	67.03	Crash Service	1
2	6/13/18 0:00	2018	6	13	0	2	0	30.23406	-97.86478		clear-night	0	9.997	0.78	5.89	64	31	2	-22.23618899	78.17	70.85	Crash Urgent	1
3	6/13/18 19:00	2018	6	13	19	2	0	30.237	-97.82774		clear-day	0	9.997	0.39	6.4	64	31	2	17.94198855	90.94	62.3	Crash Urgent	1
4	6/13/18 22:00	2018	6	13	22	2	0	30.20733	-97.83557		clear-night	0	9.997	0.52	8.21	64	31	2	-16.30179008	84.3	64.74	Crash Urgent	1
5	6/13/18 22:00	2018	6	13	22	2	0	30.237	-97.82774		clear-night	0	9.997	0.52	8.21	64	31	2	-16.28777952	84.3	64.74	Crash Urgent	1
6	6/13/18 12:00	2018	6	13	12	2	0	30.22143	-97.83638		partly-cloudy-day	0	9.997	0.47	6.32	64	31	2	68.48007876	90.65	67.82	Crash Service	1
7	6/13/18 23:00	2018	6	13	23	2	0	30.20733	-97.83557		clear-night	0	9.997	0.6	8.34	64	31	2	-25.3508998	81.57	66.46	Crash Urgent	1

Exploratory Data Analysis:

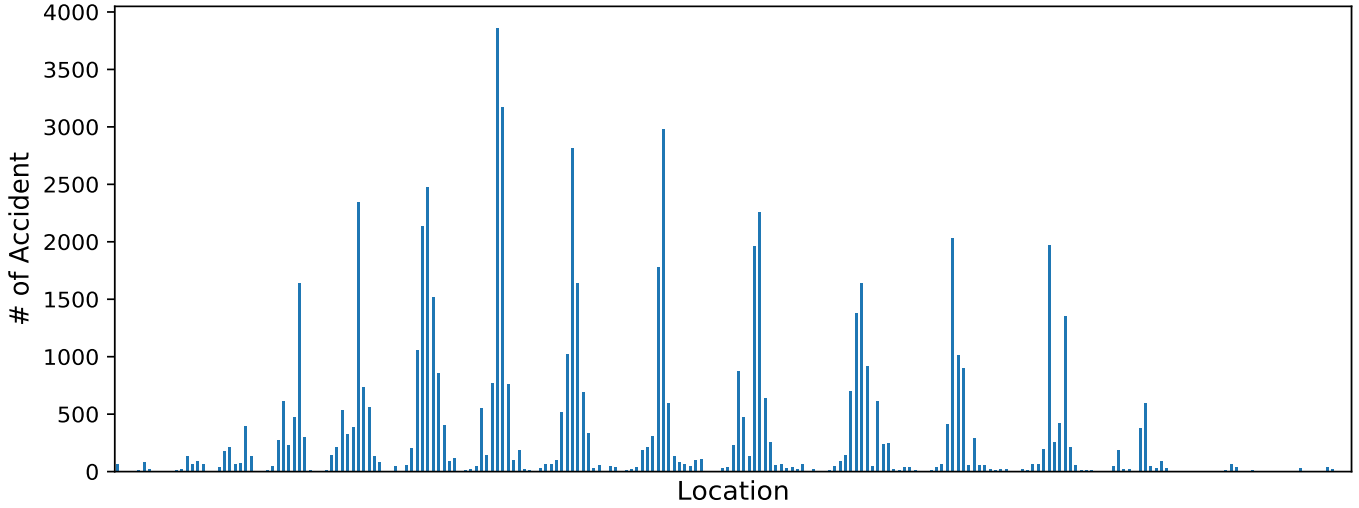
High Peaks in the Accident Curve ⟷ High Precipitation ⟷ Low Temperature



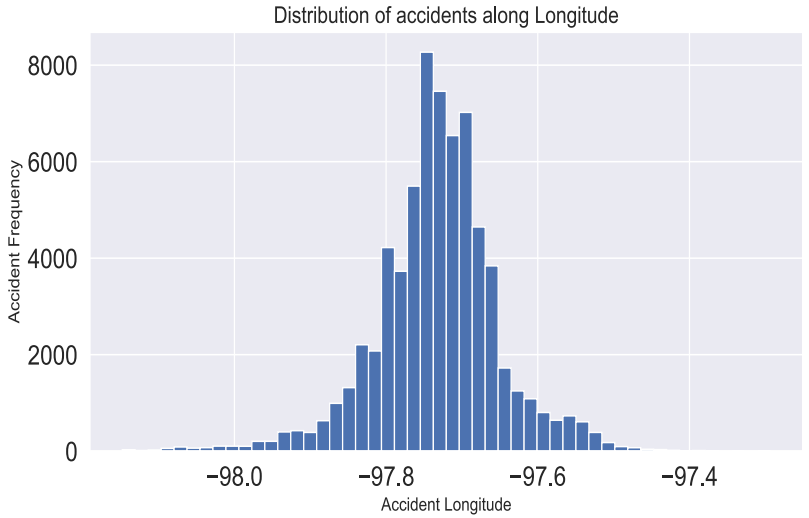
Exploratory Data Analysis:

- Independent of time features month, day, holidays
- Depends on the hour of the day
- Strongly depends on locations. Sharply dependent on geographical longitude
- We also found weak dependence of nearby traffic signals, school zones and historical landmarks

Accidents at various locations

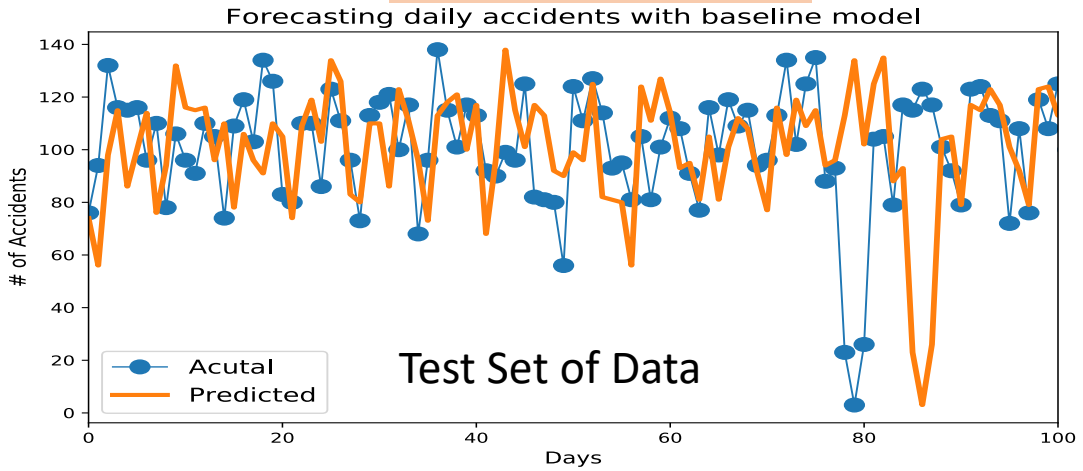


Distribution of accidents along longitude

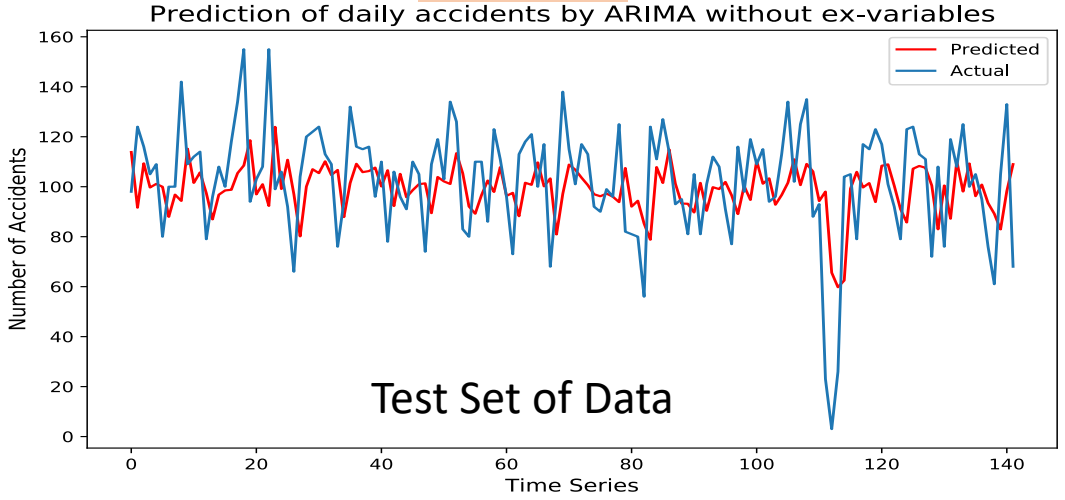


Forecasting Road Accidents in Austin: Time Series Analysis

Baseline Model



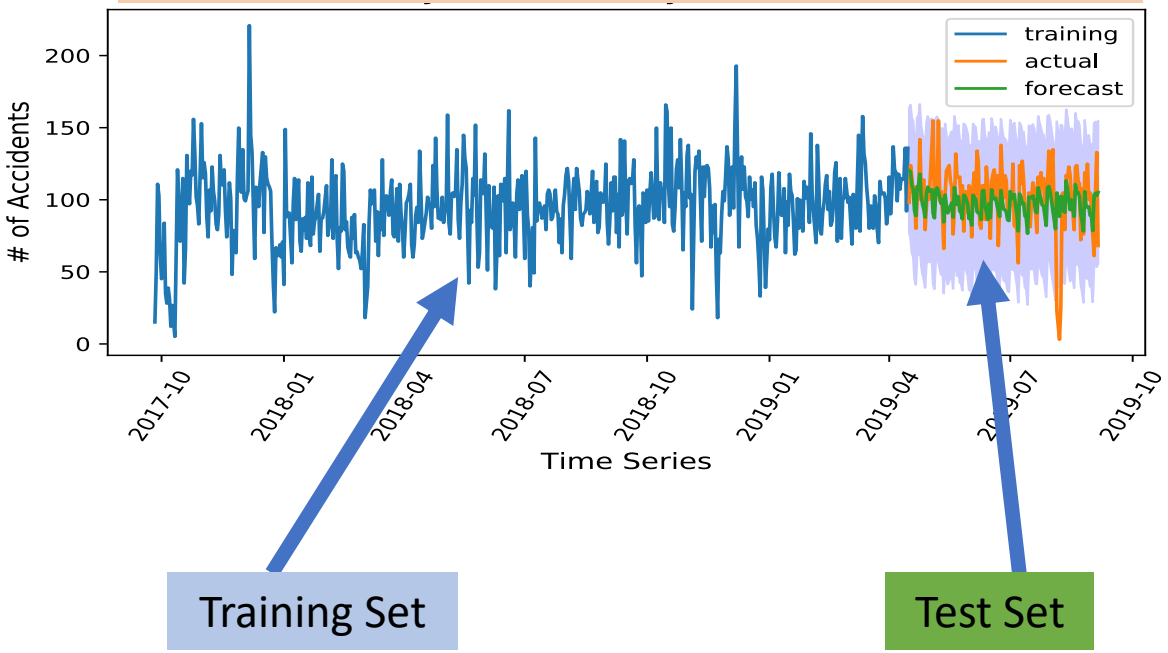
ARIMA



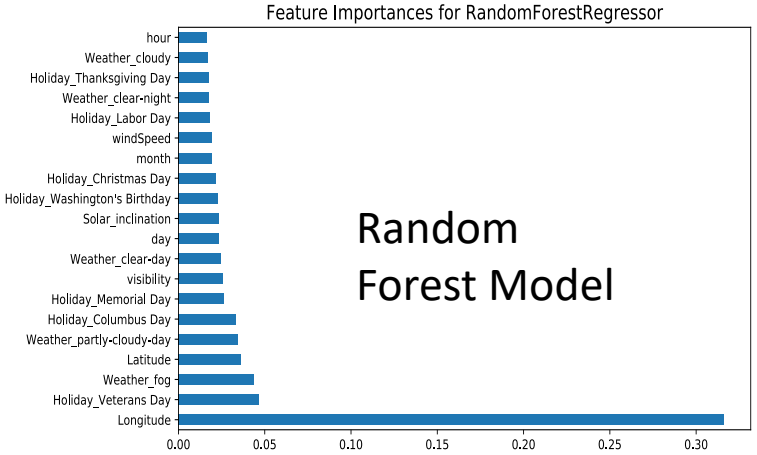
Mean absolute errors:

Baseline:	**21%**
ARIMA:	**17%**
ARIMA with exogenous variables:	**15%**

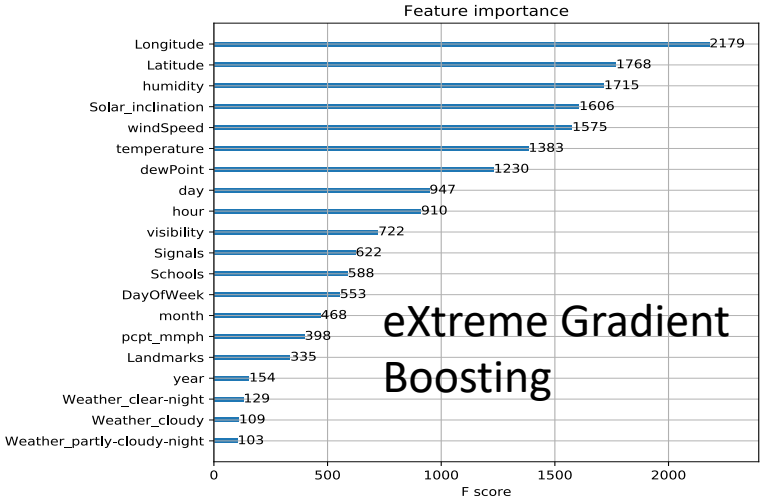
ARIMA with Exogenous variables

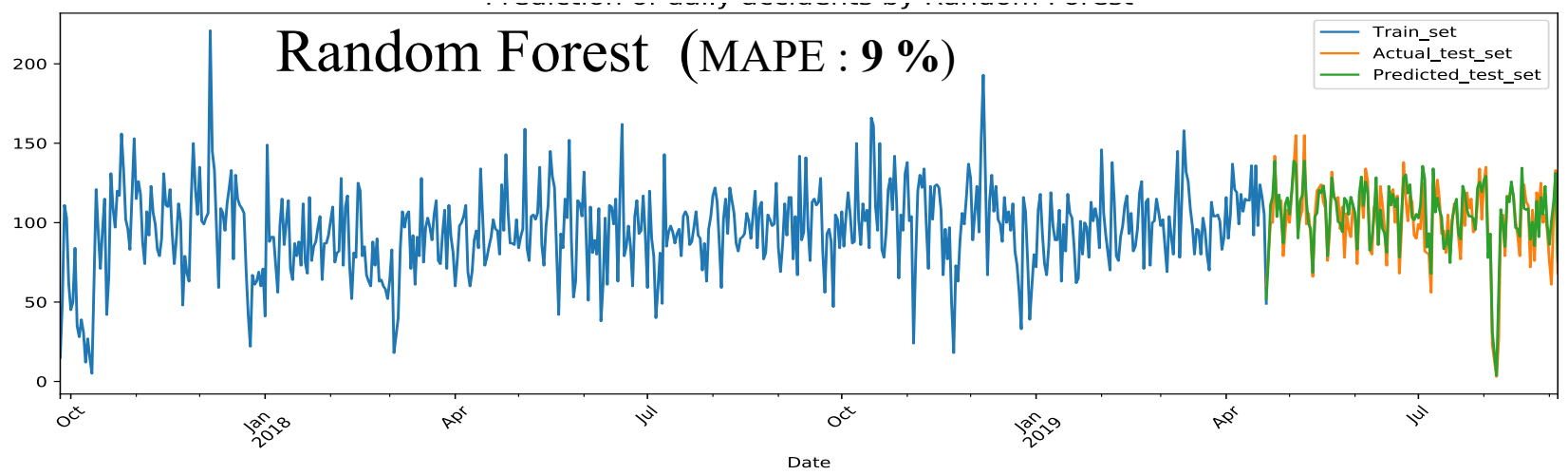


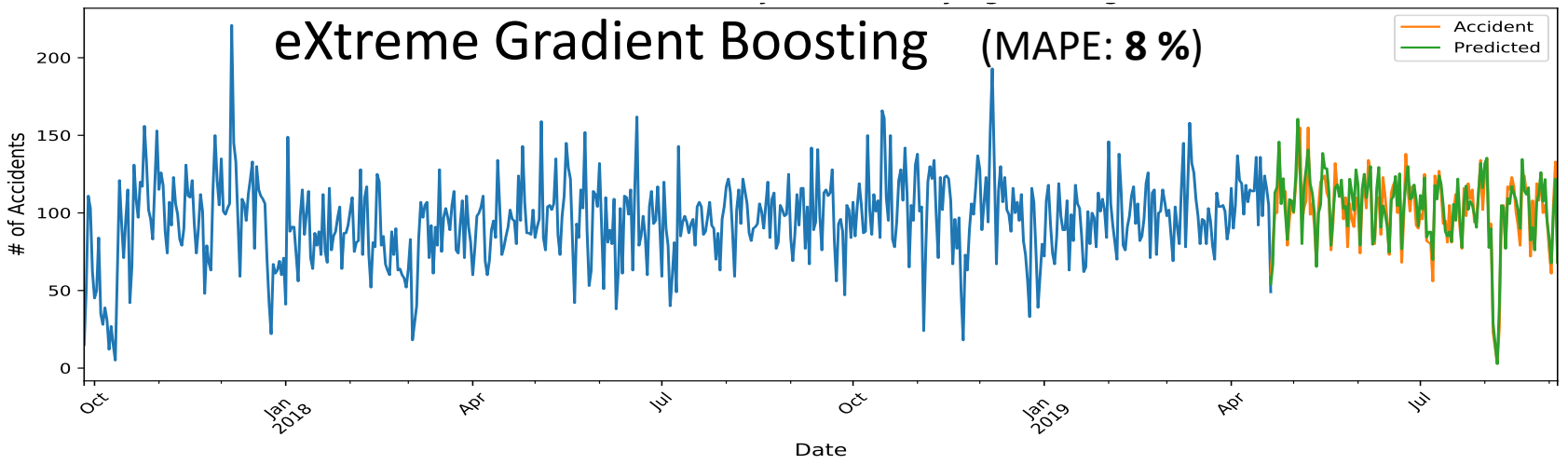
Performance of Two Decision Based Predictive Models: Daily Road Accidents in the City of Austin



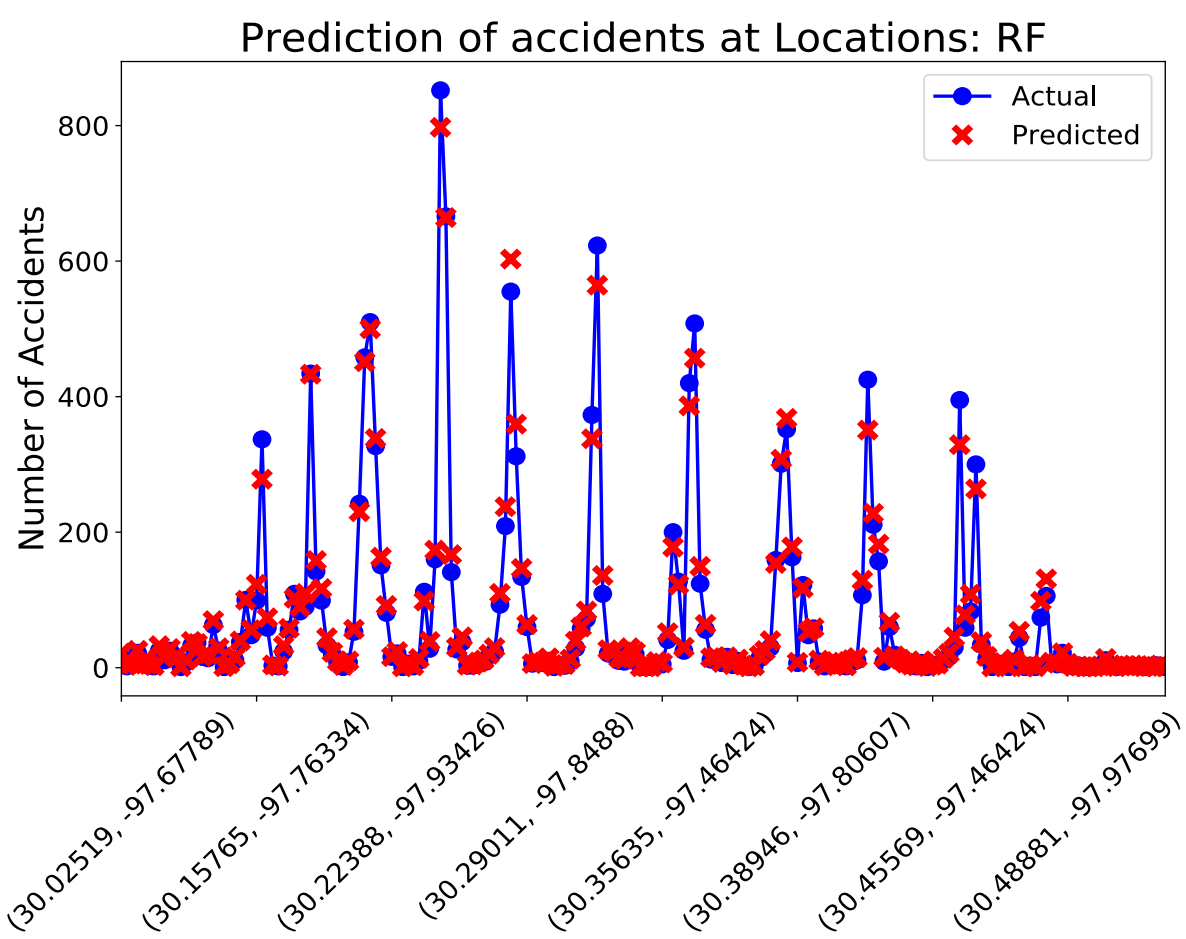
Longitude is the most Important



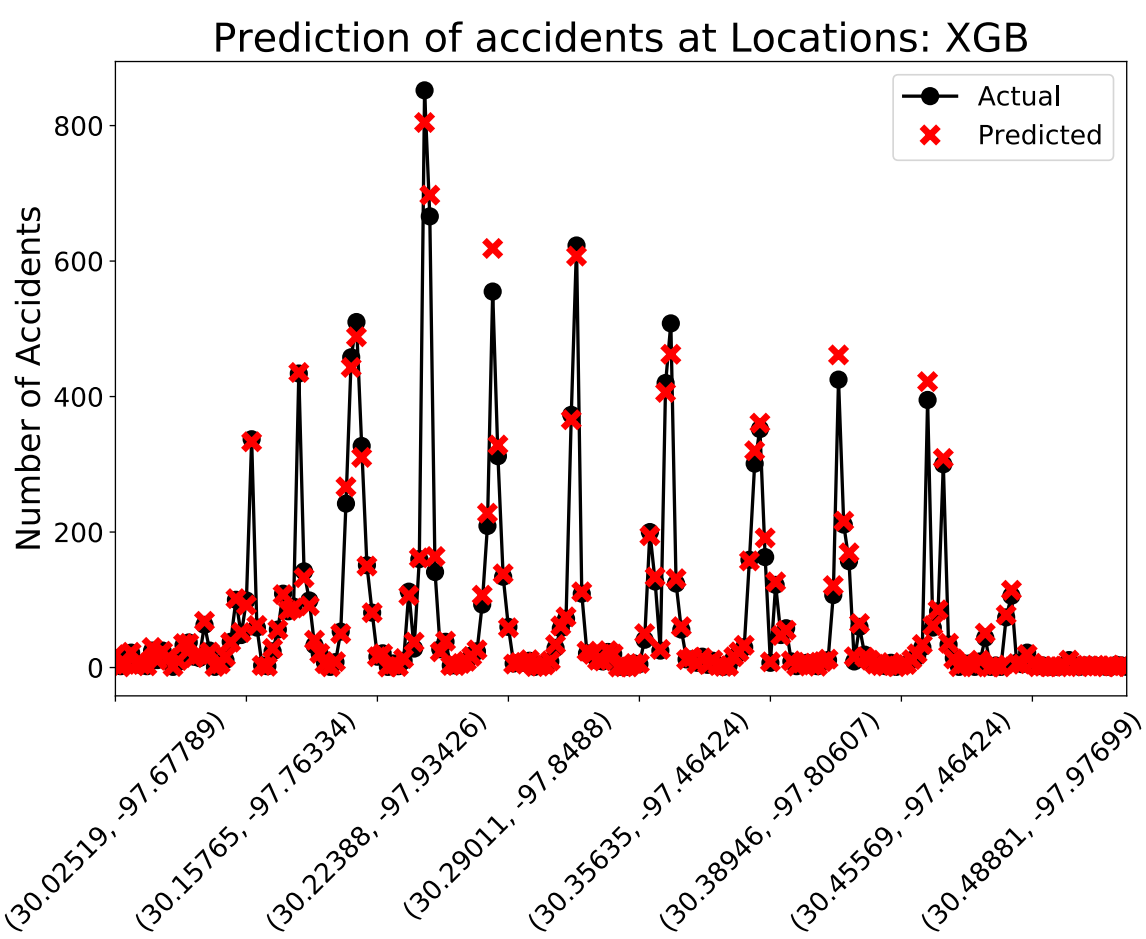




Performance of Two Decision Based Predictive Models: Road Accidents at Locations (*Latitude, Longitude*):

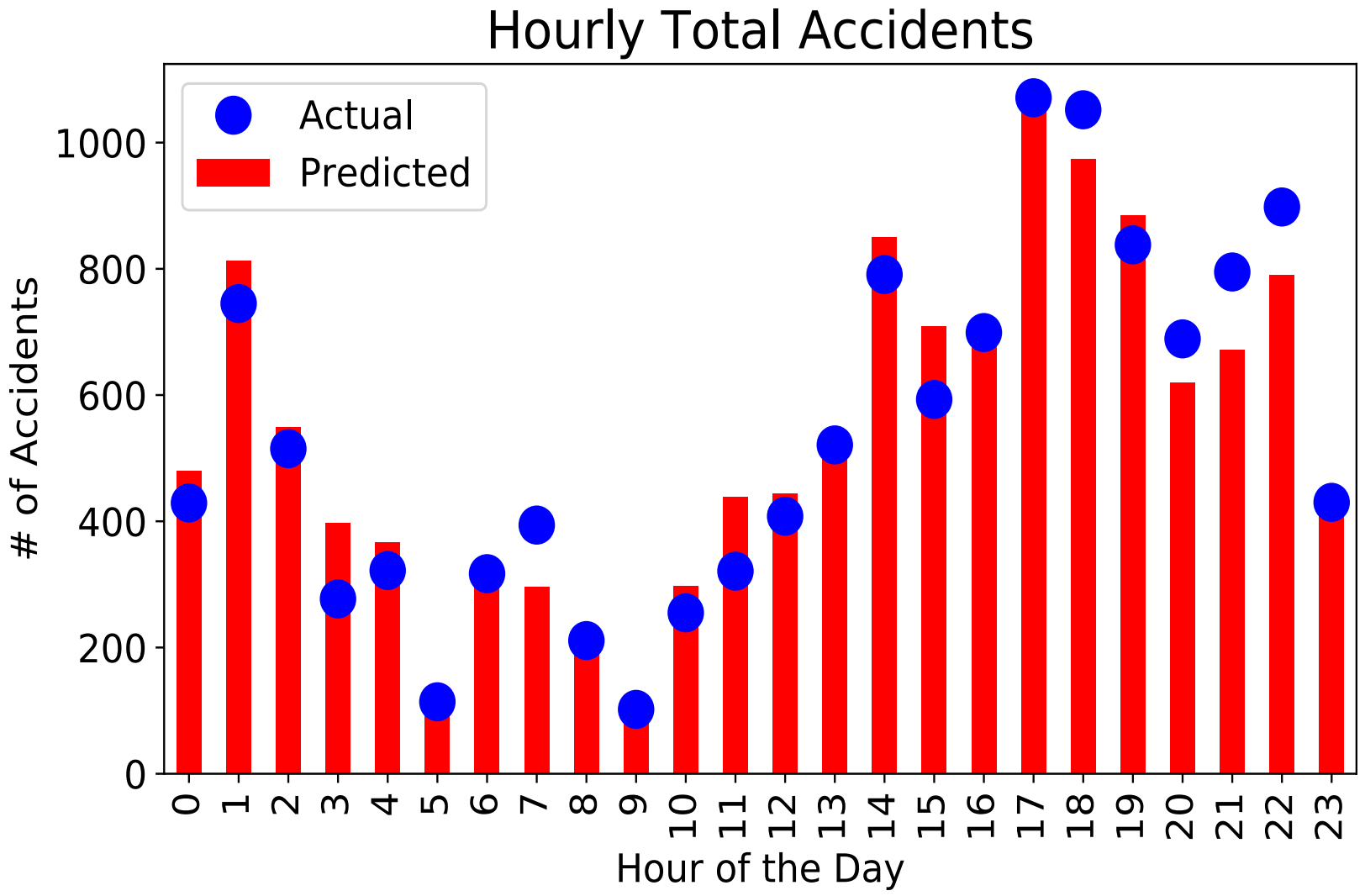


Random Forest (MAPE : **11 %**)



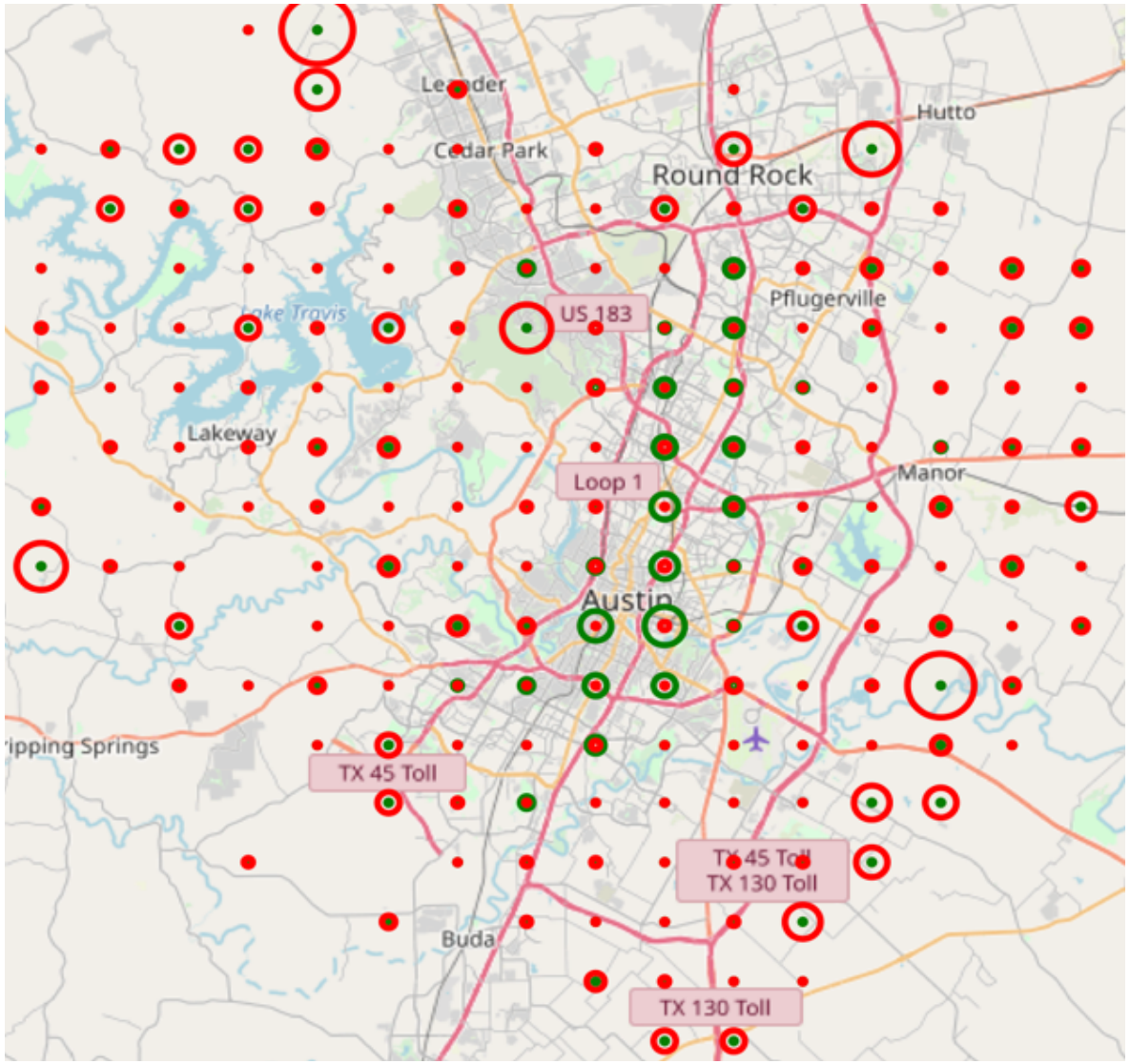
eXtreme Gradient Boosting (MAPE: **6 %**)

eXtreme Gradient Boosting Model: Prediction of Hourly Average Road Accidents



eXtreme Gradient Boosting (MAPE: **10 %**)

Locations and Prediction Errors:



Summary:

We predicted the road accidents in the city of Austin at various locations and time

Extreme gradient boosting model outperforms the other models that were used in this project.

Random forest model performs far better than the autoregressive time series model ARIMA, although ARIMA itself performed better than the baseline model for time series analysis.

Location wise accidents have been predicted with performance better than the daily accident counts.