

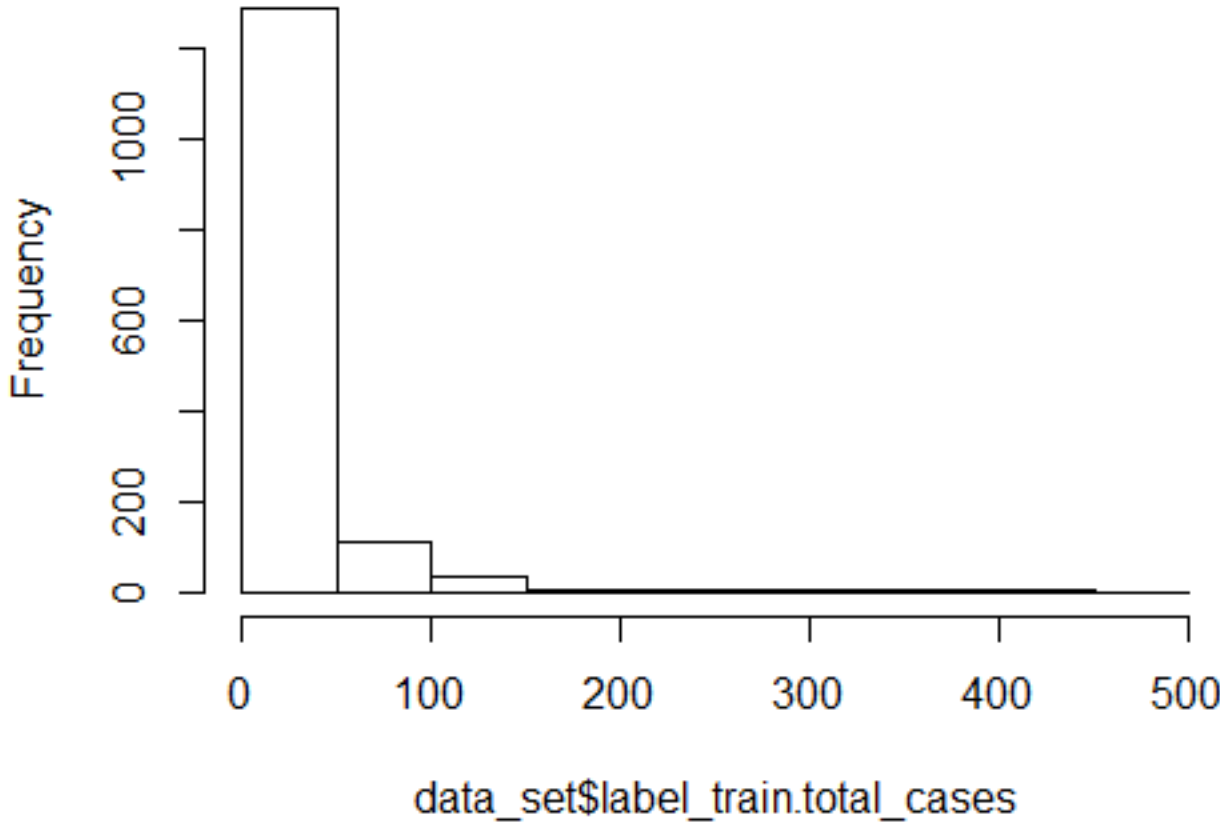
BUSINESS PROBLEM

Dengue fever is an epidemic spreading in tropical and subtropical areas. The sever cases of the this disease would cause severe bleeding beyond fever, rash, and muscle with joint pain. There are no vaccines to prevent it. The only thing we can do is to avoid mosquito bites. We analyzed the dataset of San Juan and Iquitos to predict how the climate change influences the disease spread.

DATA INSIGHTS

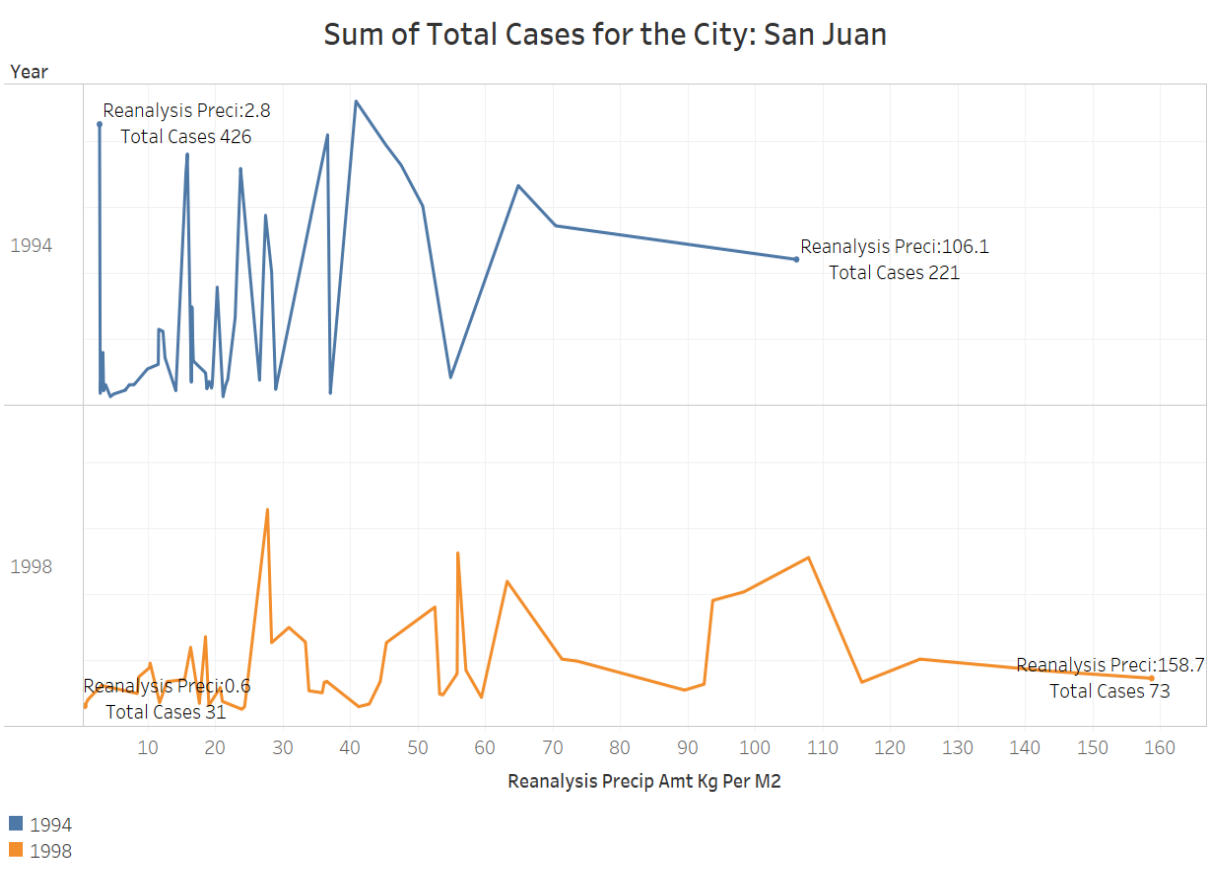
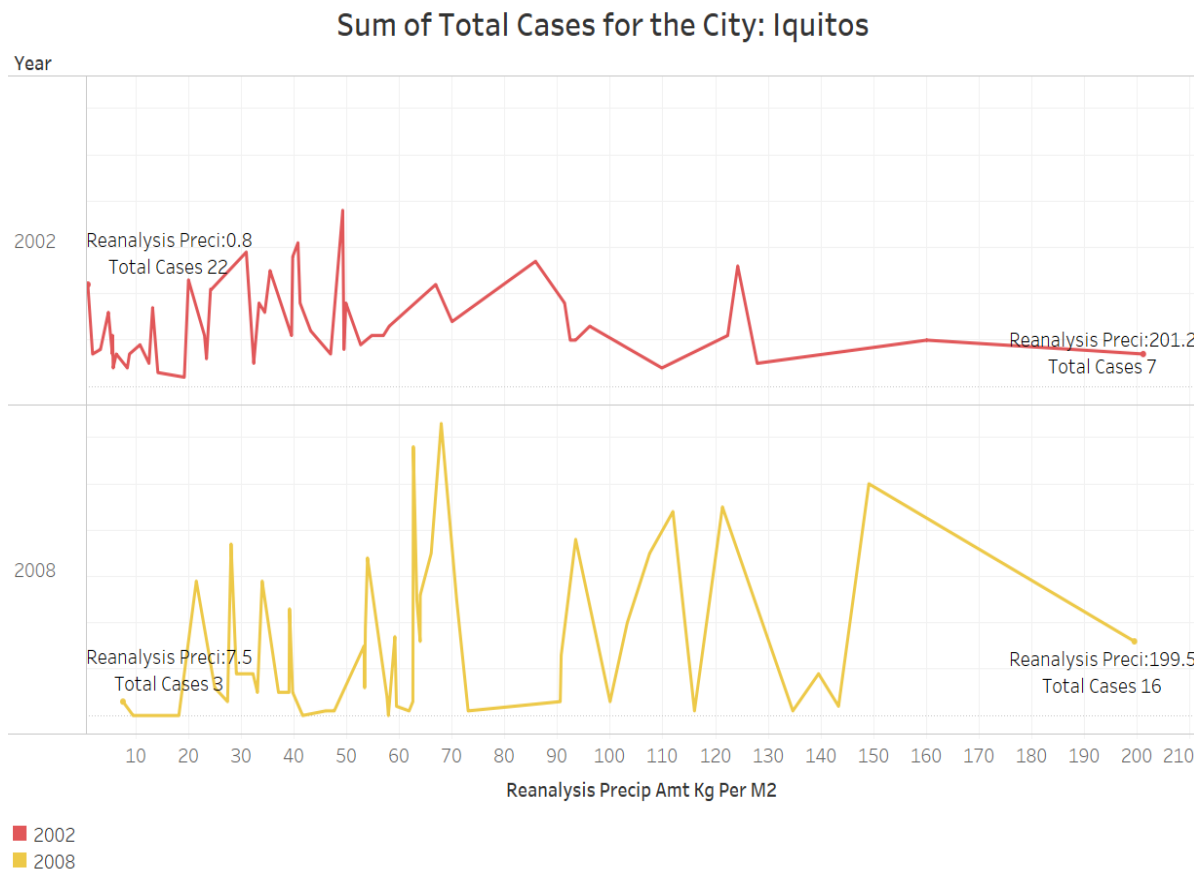
We leveraged KNN, Decision Tree, Association Rules, SVM to analyze the dataset. After analysis, we found that a similar trend happened in two cities in specific climate condition.

Histogram of data_set\$label_train.total_cases



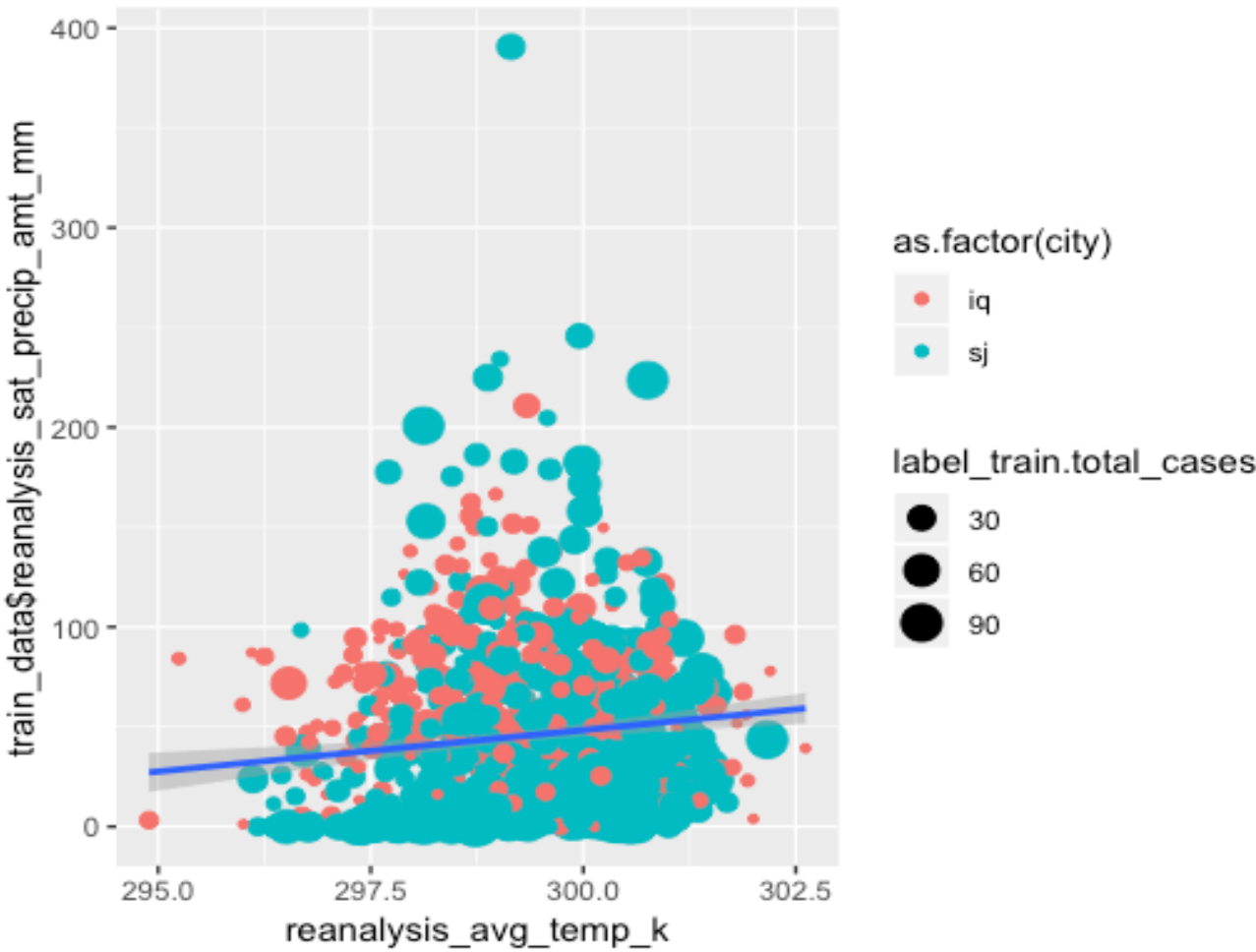
PRECIPITATION

The precipitation in a specific range causes the burst of Dengue Fever in the two cities. Especially a small precipitation, it leads Dengue Fever spread.



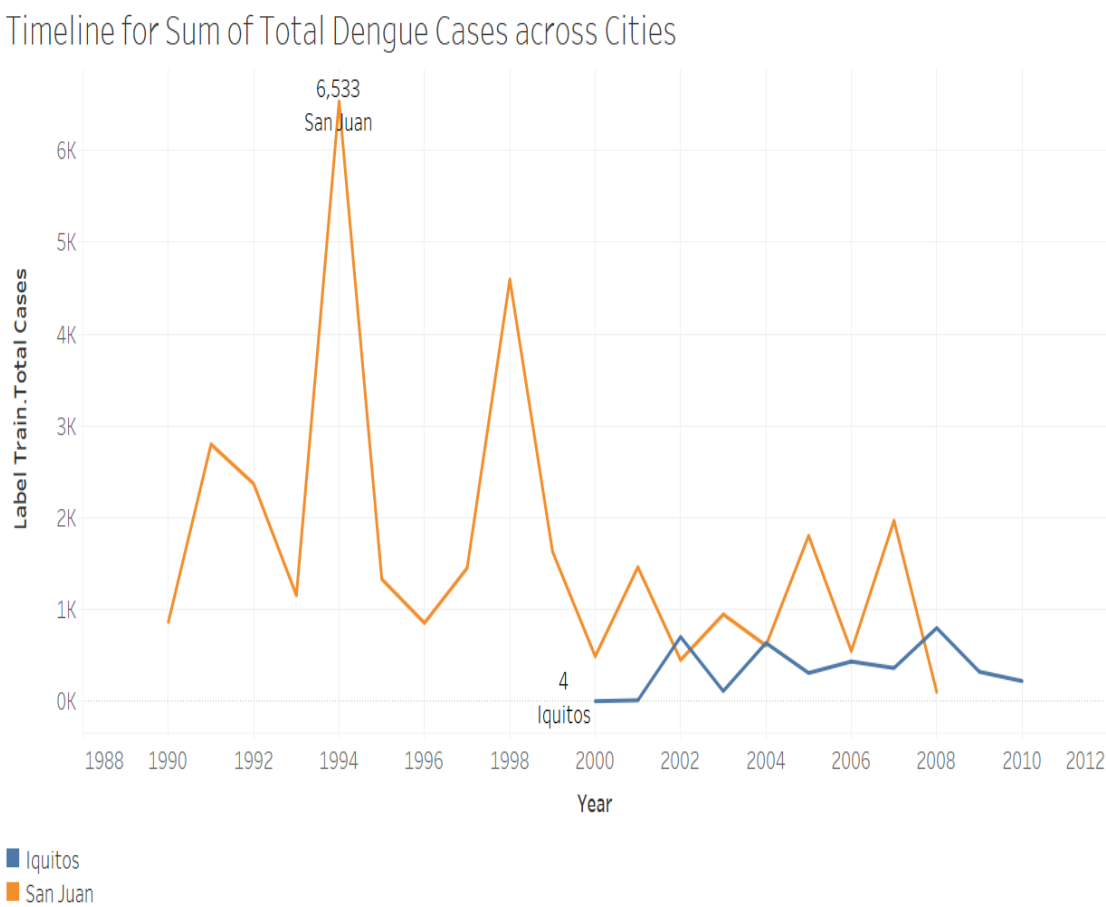
TEMPERATURE

Temperature is the other reason for Dengue Fever spread. The higher the temperature, the more Dengue Fever cases.



CITY: SAN JUAN IQUITOS

These two cities are all at tropical or subtropical area. San Juan has long term Dengue Fever problem and has more cases.



DATA MODELS

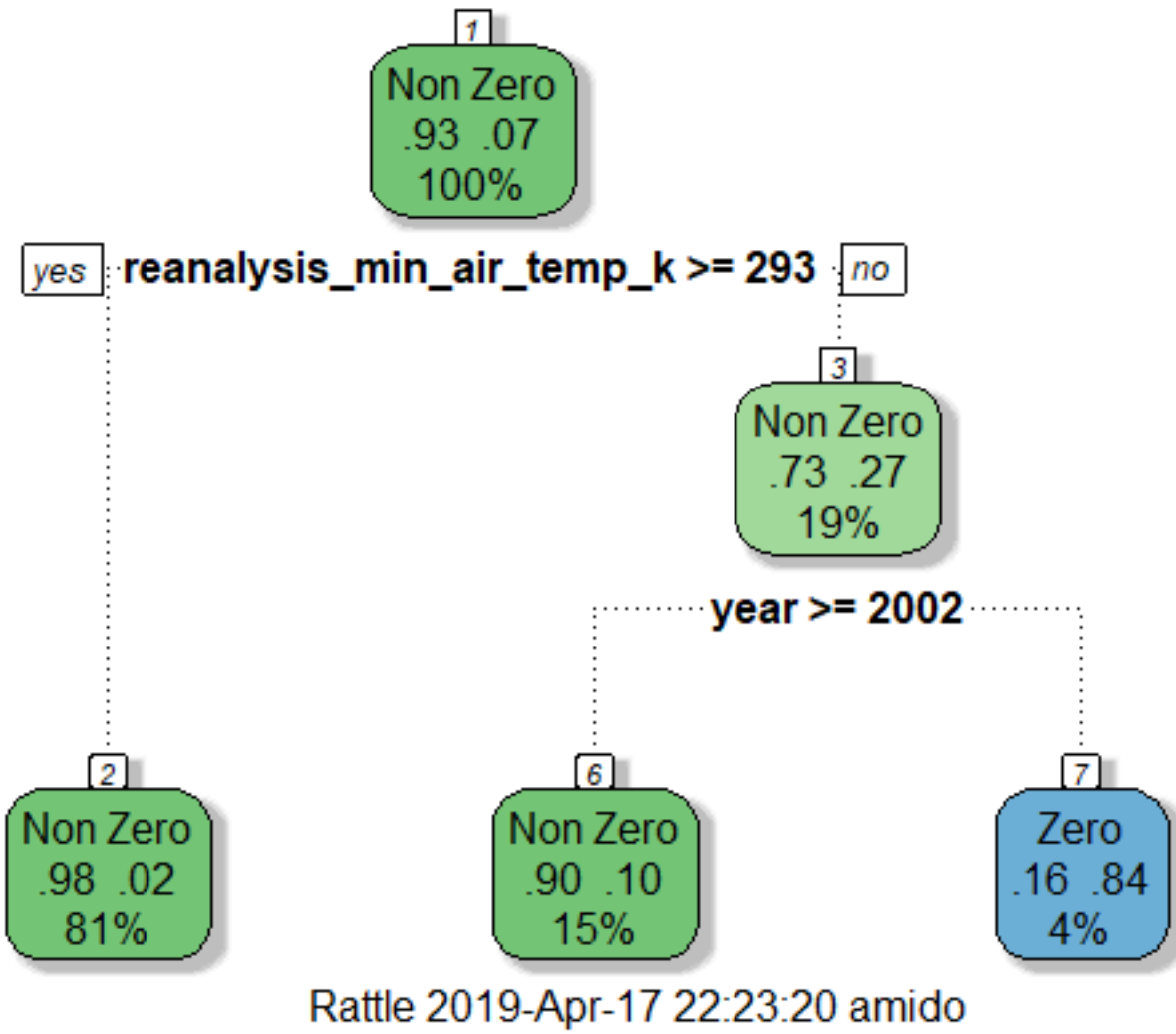
We leveraged KNN, Decision Tree, Association Rules, Naïve Bayes, Random Forest, SVM to gain insights.

ASSOCIATION RULES

- We use association rules to find relationships between the climate parameters and total dengue cases. We found some interesting relations stating:
- The San Juan city has higher total dengue fever cases with exceptionally high cases in the year 1994.
 - High temperature leads higher dengue fever cases.
 - High precipitation leads higher dengue fever cases.

DECISION TREE

We noticed that there are zero total cases in the dataset. So we use decision tree to find parameters leading to the same. Our observation: The cases with minimum air temperature less than 293 and occurring before the year 2002 have a probability of 84% to be a record of Zero total cases.



NAIVE BAYES

The accuracy for the Naive Bayes model was 88% which is too low. Thus we wouldn't be considering it .

Confusion Matrix and Statistics			
Reference			
Prediction	low	middle	high
low	256	31	3
middle	0	0	0
high	0	0	0

SVM, RANDOM FOREST, BAGGING, KNN, XGB, GBM

We predict the number of total cases from year 2008 to 2013 using models: SVM, Random Forest, Bagging, KNN, XGB and GBM, each giving different accuracies close to 90% and above. Amongst them, KNN shows the least accuracy while XGB showing the most.

