FINAL PROJECT SEGMENT 2

GROUP 7

PRESENTATION

Selected Topics

Data Sources

Data exploration

Data Analysis

Dashboard

Selected Topic:

AMBULANCE PATIENT OFFLOAD TIMES

Analysis of the Ambulance Patient Offload Times (APOT) in Sacramento County. The goal of this project is to identify what is influencing the extended wait times. Shedding a light on where the problems are will show the parties involved where to look and help initiate a plan to alleviate the current situation. The questions we hope to answer are:

1. Which impressions has the longest/shortest APOTs? (Overall and by Hospital codes)
2. Compare impressions to zip codes. Which zip codes have the highest/lowest rates of which impressions?
3. Which zip codes utilize ambulance transports to the ER the most in 2017-2022?
4. Wait times per hospital by date/time
5. Do any patterns arise from the data?

DATA SOURCES

The data selected for this project was obtained from Sacramento County Emergency Medical Services Agency. The data set includes the following data points :

- Destination Hospital identifier
- Destination Hospital location Latitude and Longitude
- Date/Time of Occurrence
- Ambulance Patient Offload Time (APOT)
- Patient Complaint identifier
- Ambulance Agency Identifier
- Ambulance Unit identifier
- Postal Code of Where call was Originated

DATA EXPLORATION

STRATIFICATION

Through machine learning it was found that there is a strong relationship between APOT and hospital codes. To further explore the data and understand this relationship data is being separated into subgroups of individual hospitals as part of the stratification process.

SCIKIT LEARN

- Linear Regression
- Balanced Random Forest Classifier
- Easy Ensemble AdaBoost Classifier
- Oversampling
- SMOTE Oversampling
- Under sampling
- Combination (Over and Under) Sampling

TRAINING/TESTING SETS

The testing and training sets are created as Status. The code used for this is below:

- X = df.drop(columns='Status')
- X = pd.get_dummies(X)
- *# Create our target*
- y = df[target]

DATA ANALYSIS

PROCESS

Data extracted and cleaned, Machine Model created, algorithm used included:

1. Linear Regression
2. Balanced Random Forest Classifier
3. Easy Ensemble AdaBoost Classifier
4. Oversampling
5. SMOTE Oversampling
6. Undersampling
7. Combination (Over and Under) Sampling

The imbalanced-learn and scikit-learn libraries were used to evaluate data and resampling and oversampling and smote algorithms the data set is resampled to acquire the count of target classes, train logistic regression classifier and calculate the balanced accuracy score as well as generate a confusion matrix and generate a classification report.

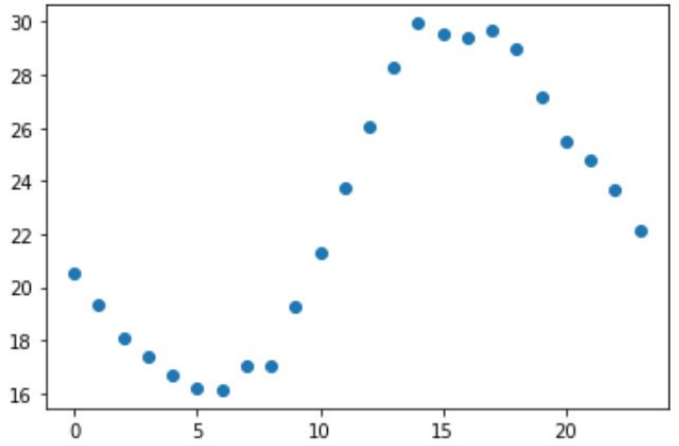
Template Outline

Ambulance Patient Offload Times

Confusion Matrix

Predicted	primary Impression	true	

APOT Average per Hour



Ranking (Random Forest)

Importance of Features	Featured