IST 707: COVID-19 PATIENT PRE-CONDITION ANALYSIS

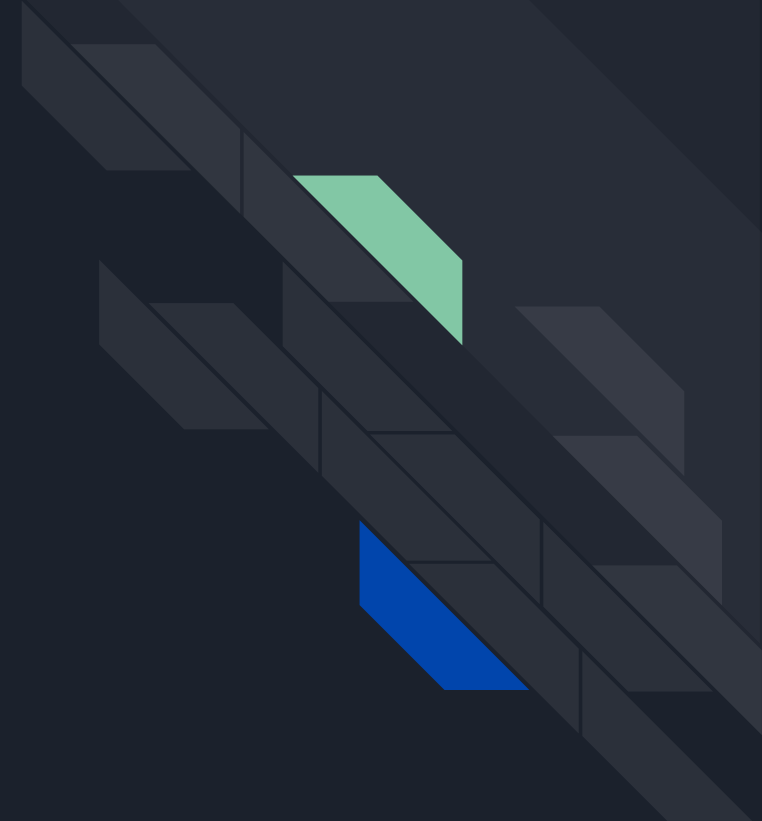ABHIJIT GOKHALE

SHUBHAM SHARMA

Syracuse University
School of Information Studies
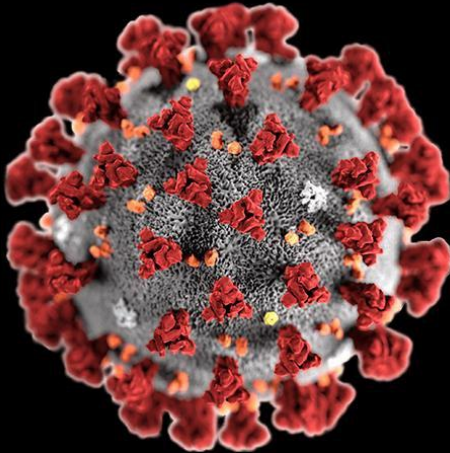
School of
Information Studies

Agenda

Why COVID Analysis?

- Provides us with an opportunity to ponder and reflect over what we can do better in the way we deal with healthcare

- Helpful in predicting what kind of resource an individual might require at the time of being tested positive

- Helpful in triage of patients

Project Objective

- **Solve the classification problem of predicting ICU requirement and analyzing its relation with mortality rate**

- **Analyzing COVID trends through data visualizations.**

Project Lifecycle

1 Data Preprocessing

Remove redundant columns and null values, Perform Visualizations and correlation

2 Feature Engineering

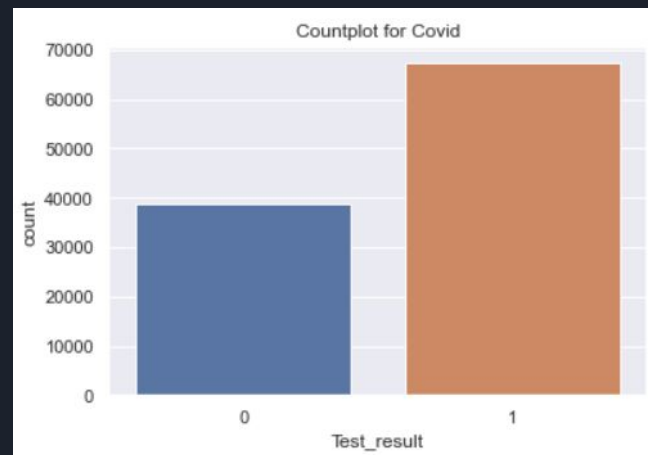Create new columns, Time series analysis for studying trend
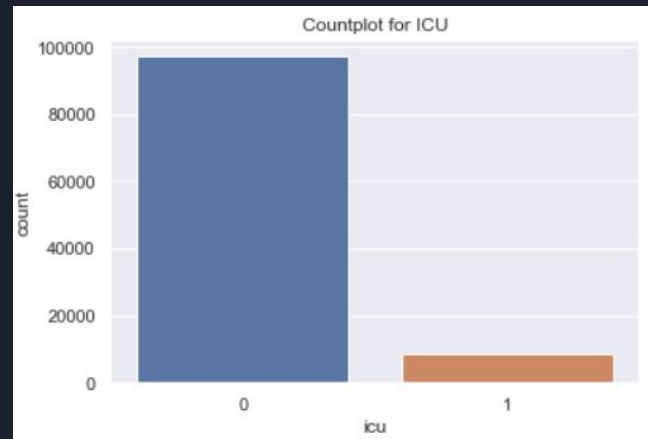
4 Model Evaluation
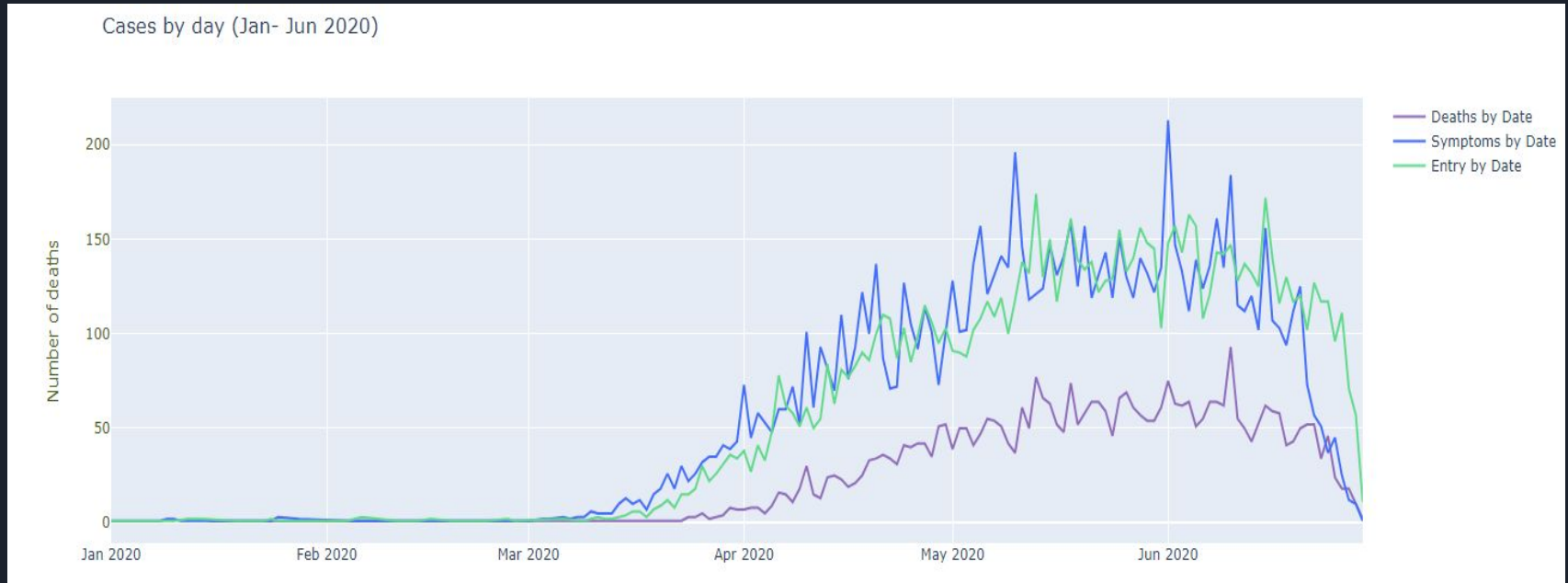
Understand performance metrics

3 Model Creation
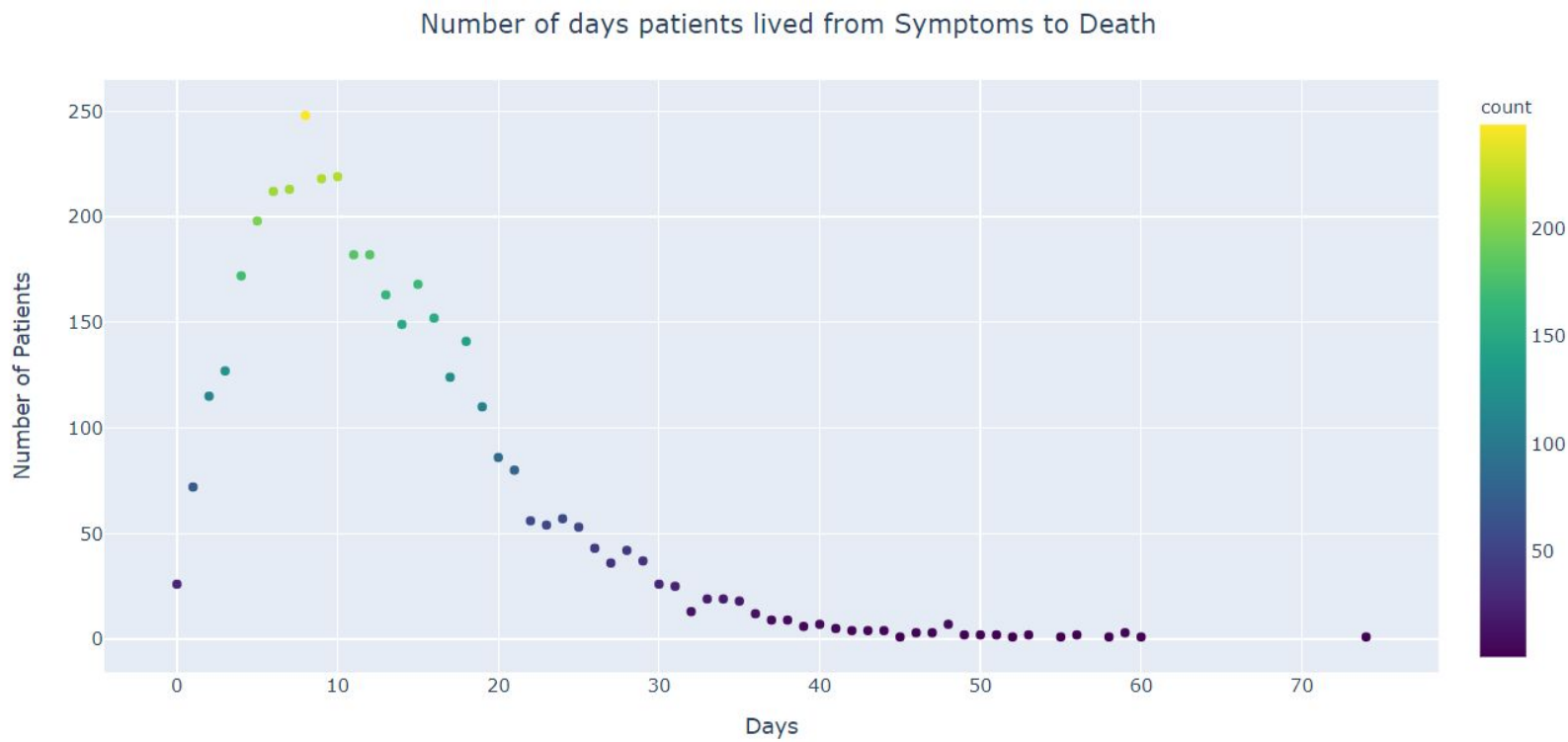
Classification Models

Understanding the Data (1)

- 23 features and 566k patient history records

- Target variable -> "icu" and "Test_result"

- Health related data of patients whether they are suffering from pneumonia, immunosuppression, diabetes, obesity, cardiovascular problems and other diseases.

- Datetime data for year 2020, telling the first onset of symptoms, date of hospitalization and date of death.

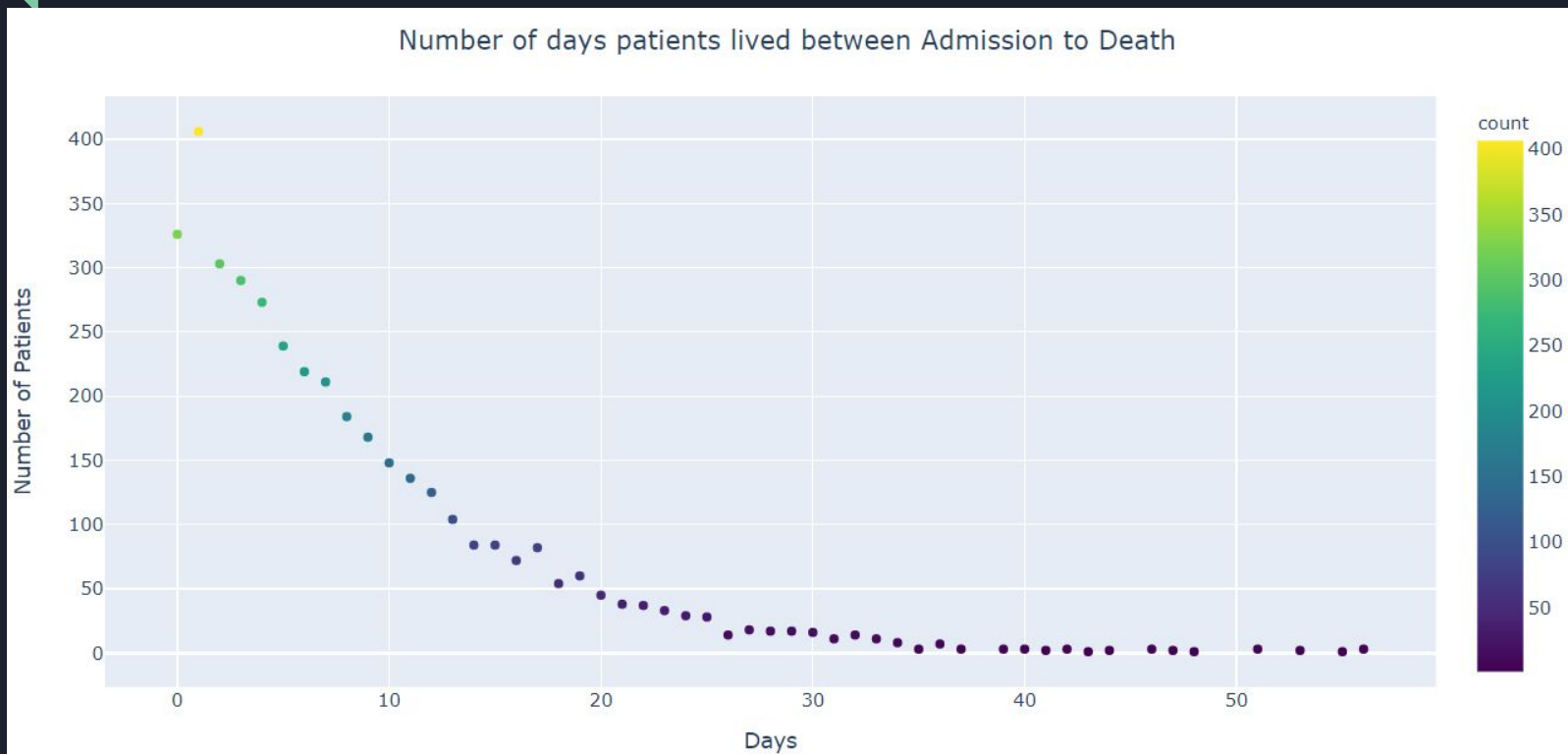# Understanding the Data: Historical Trends(1)

Cases by day (Jan- Jun 2020)

Legend:
- Deaths by Date
- Symptoms by Date
- Entry by Date

Y-axis: Number of deaths (0, 50, 100, 150, 200)

X-axis: Jan 2020, Feb 2020, Mar 2020, Apr 2020, May 2020, Jun 2020
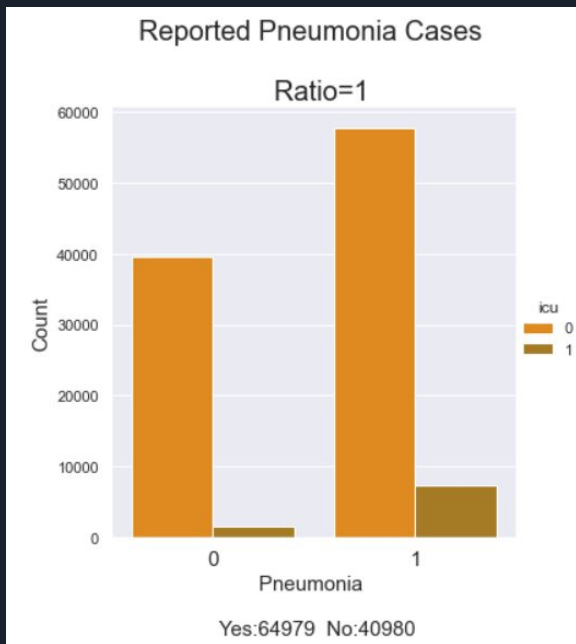
Understanding the Data: Historical Trends(2)

Number of days patients lived from Symptoms to Death

Understanding the Data: Historical Trends(3)

Number of days patients lived between Admission to Death

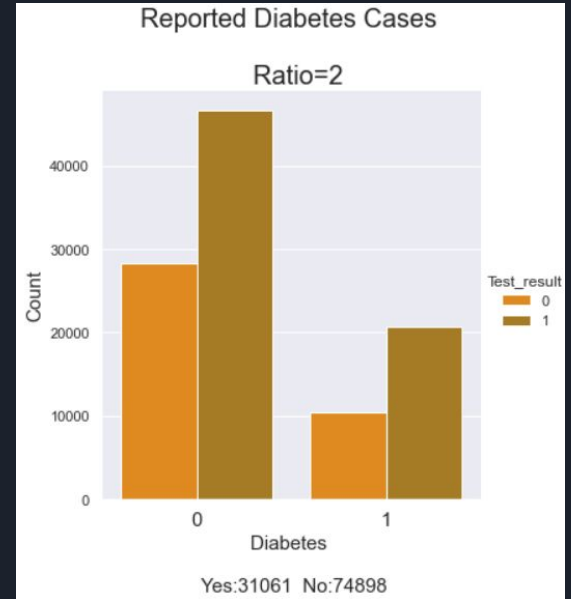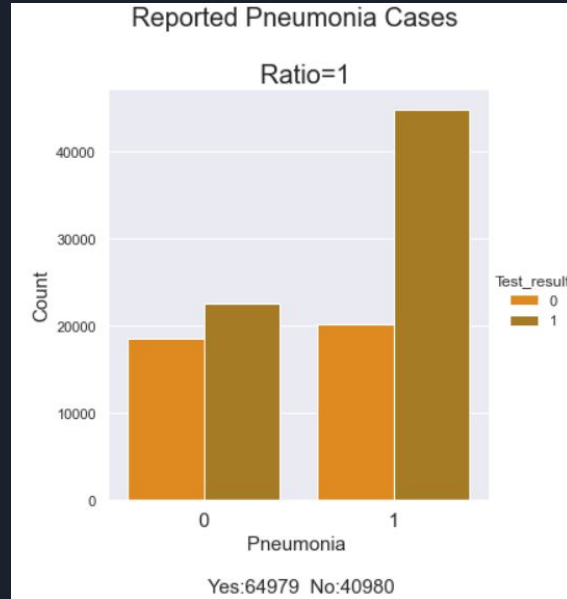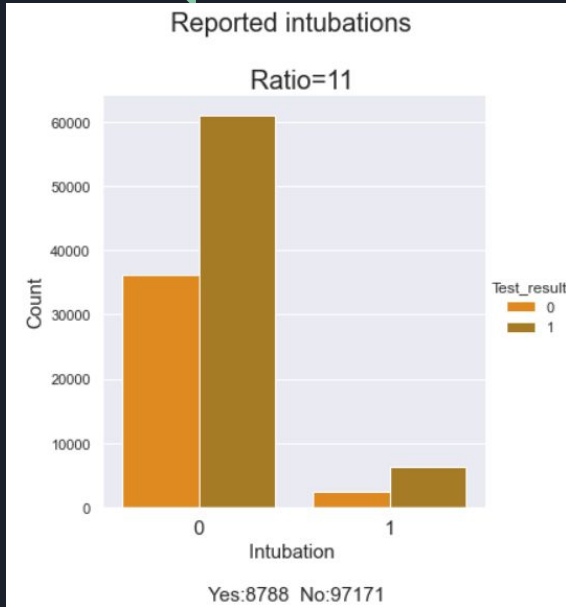# Understanding the Data: Effect On Age Groups
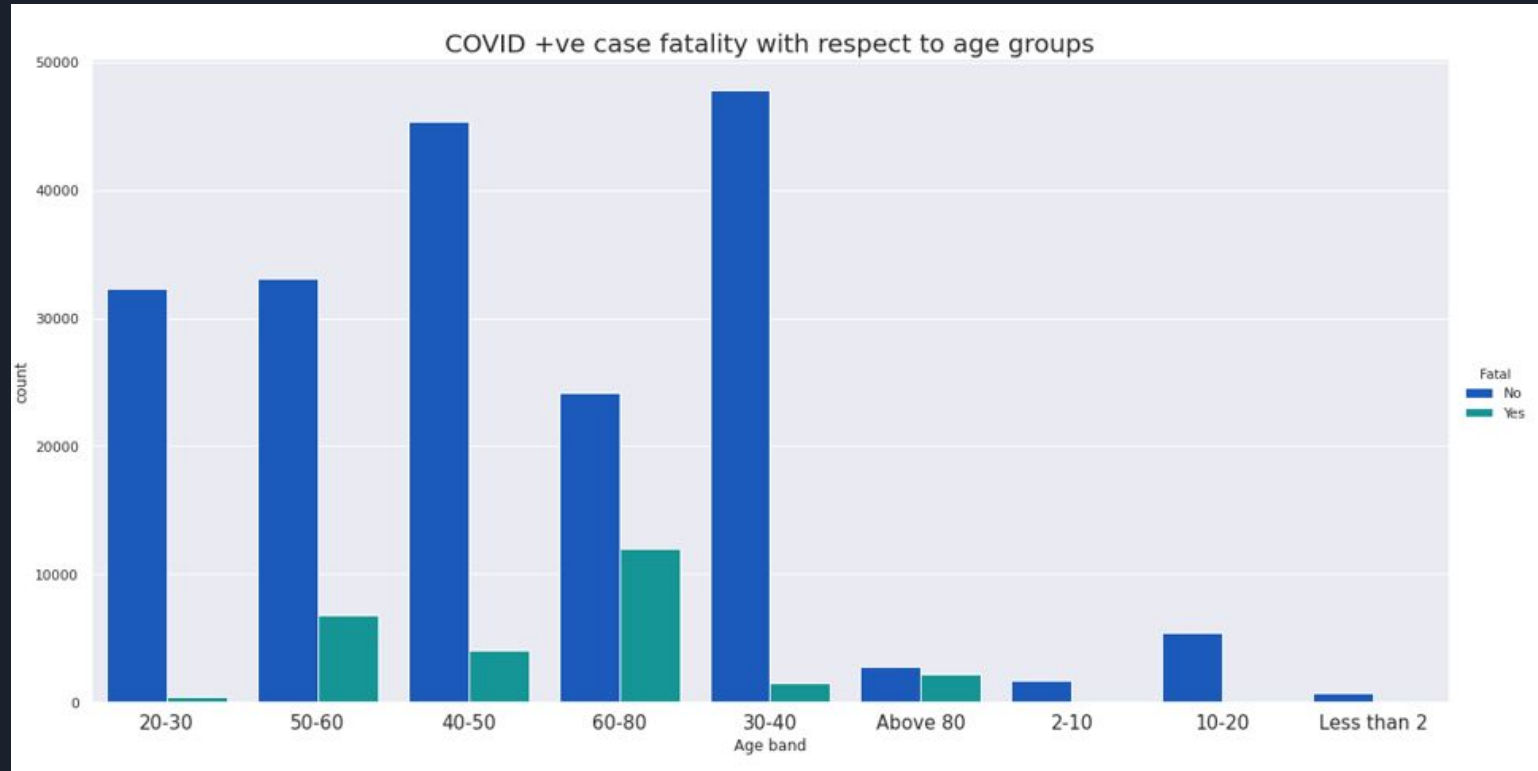
Understanding the Data: Based On Feature Importance
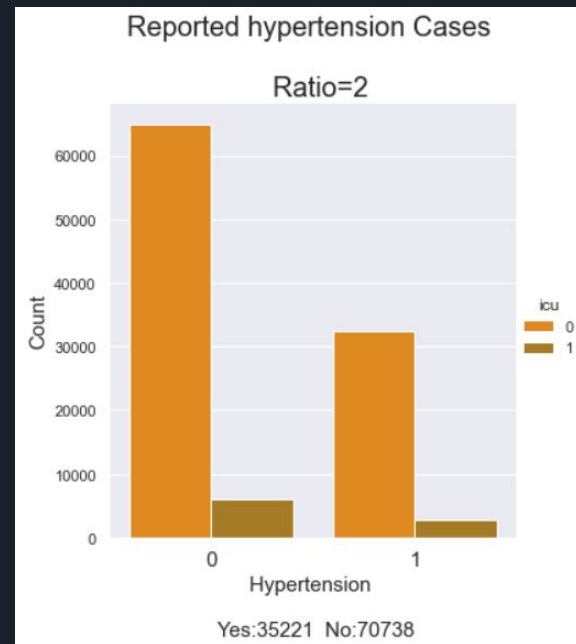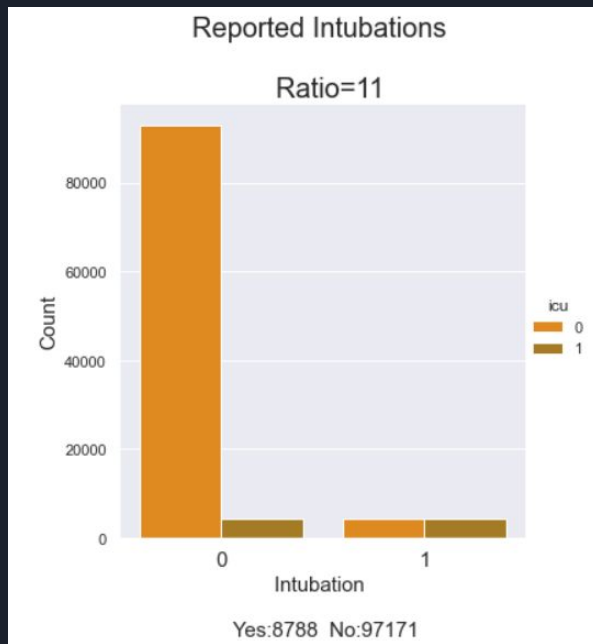
Effect of Pneumonia, Intubations and Hypertension Cases based on ICU

Understanding the Data: Based On Feature Importance

Effect of Pneumonia, Intubations and Diabetes Cases based on Covid Result

Classification Models: Predicting ICU & Covid

Logistic Regression	Decision Tree	Random Forest	Gradient Boosting
L1 and L2 penalty on the cost	Tuning Impurity, No. of trees and max tree depth	Tuning Impurity, No. of trees and max tree depth	Tuning Impurity, No. of trees and max tree depth

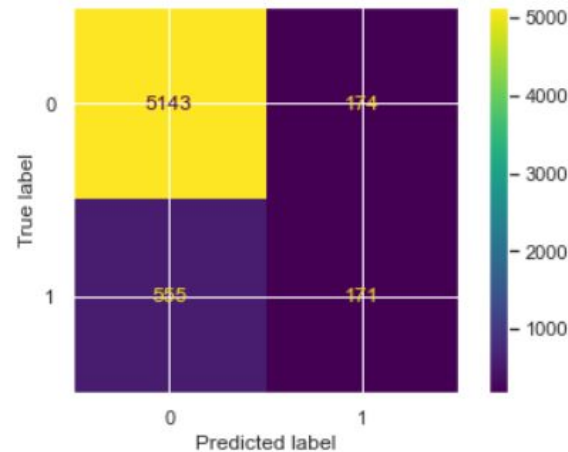$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

Did hyperparameter tuning with 3-fold Cross Validation for all models

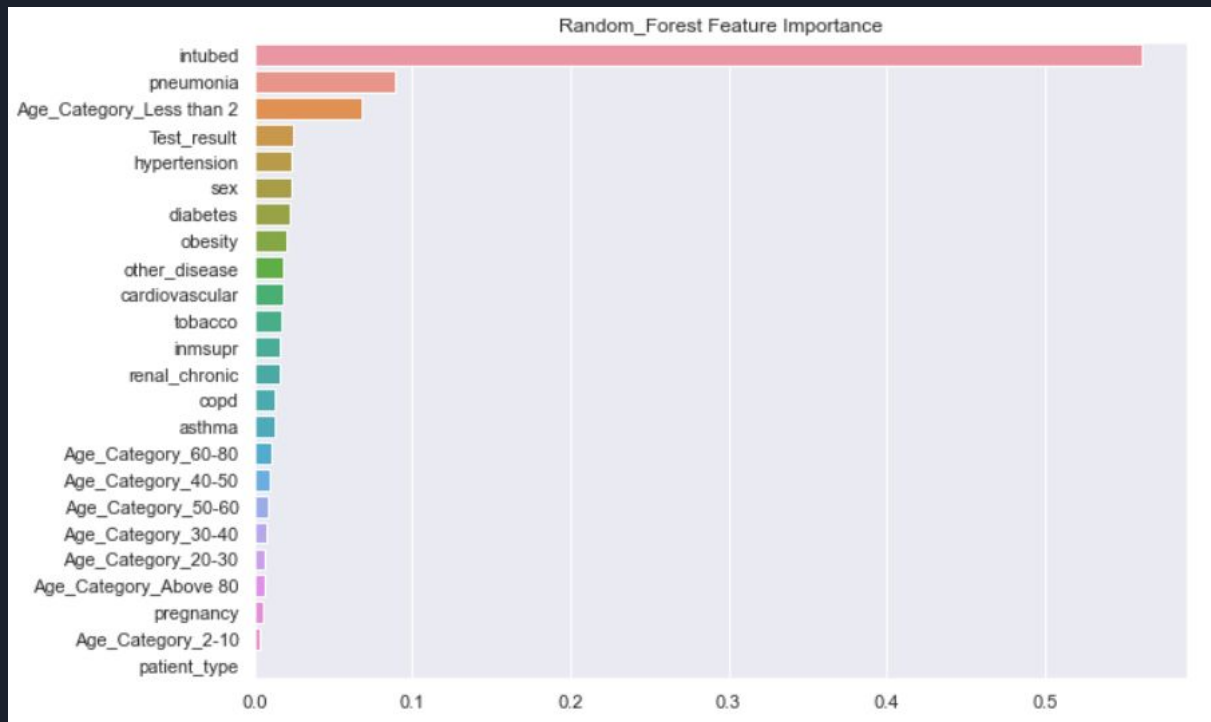# Predicting Whether Died Patients Would Have Required ICU
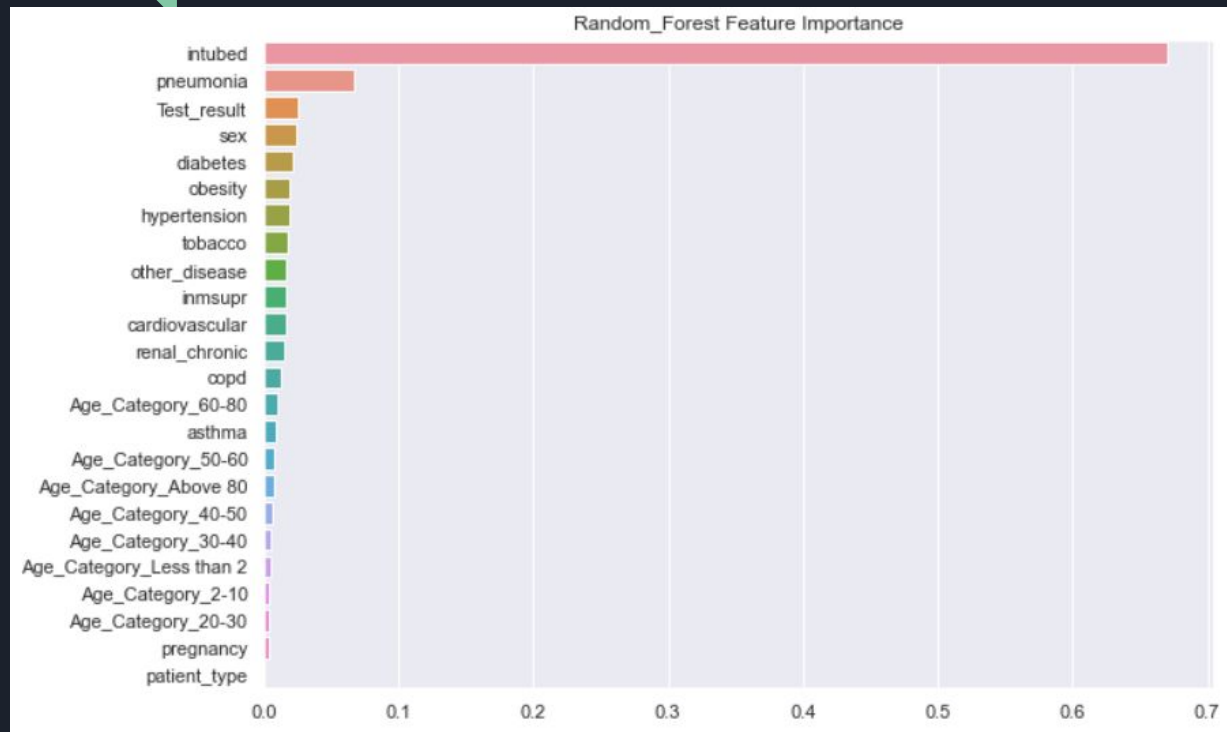
Accuracy

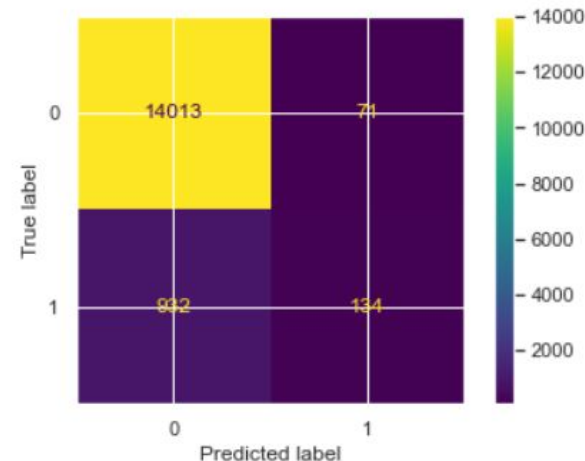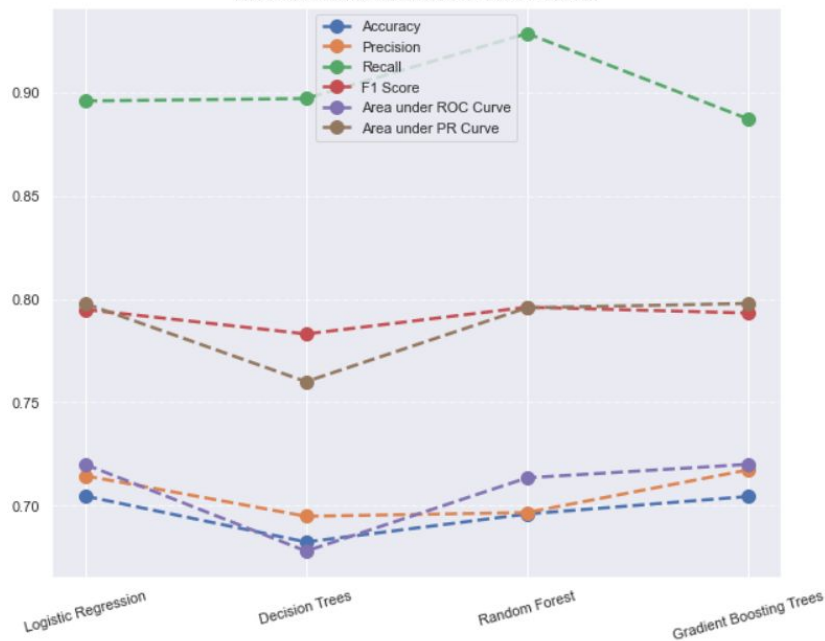# Predicting Whether Patients in Hospitalization Requires ICU

Accuracy

Predicting Whether Patient would Test Positive for COVID

Result Analysis (1)

ICU Patients who have most common preconditions
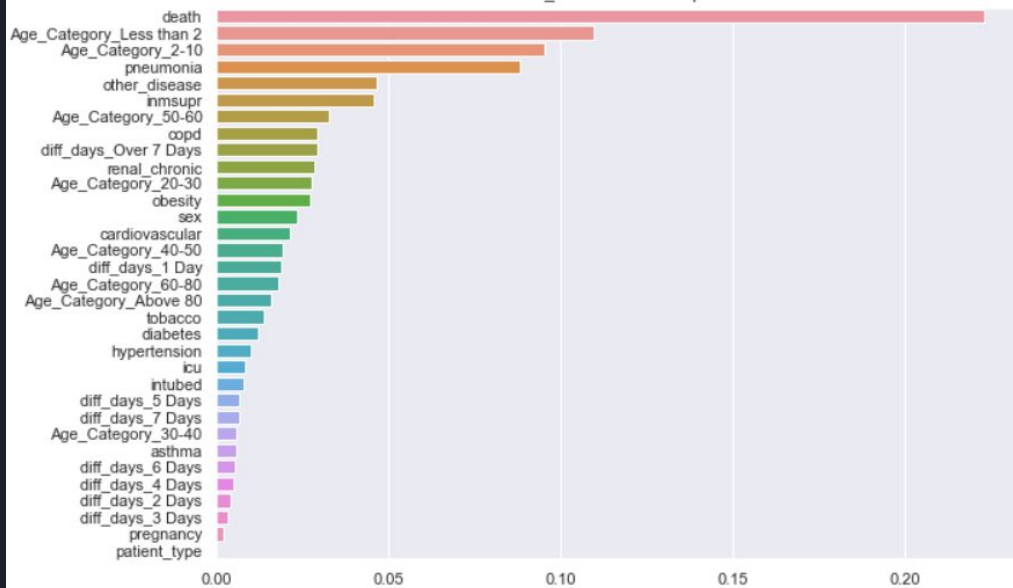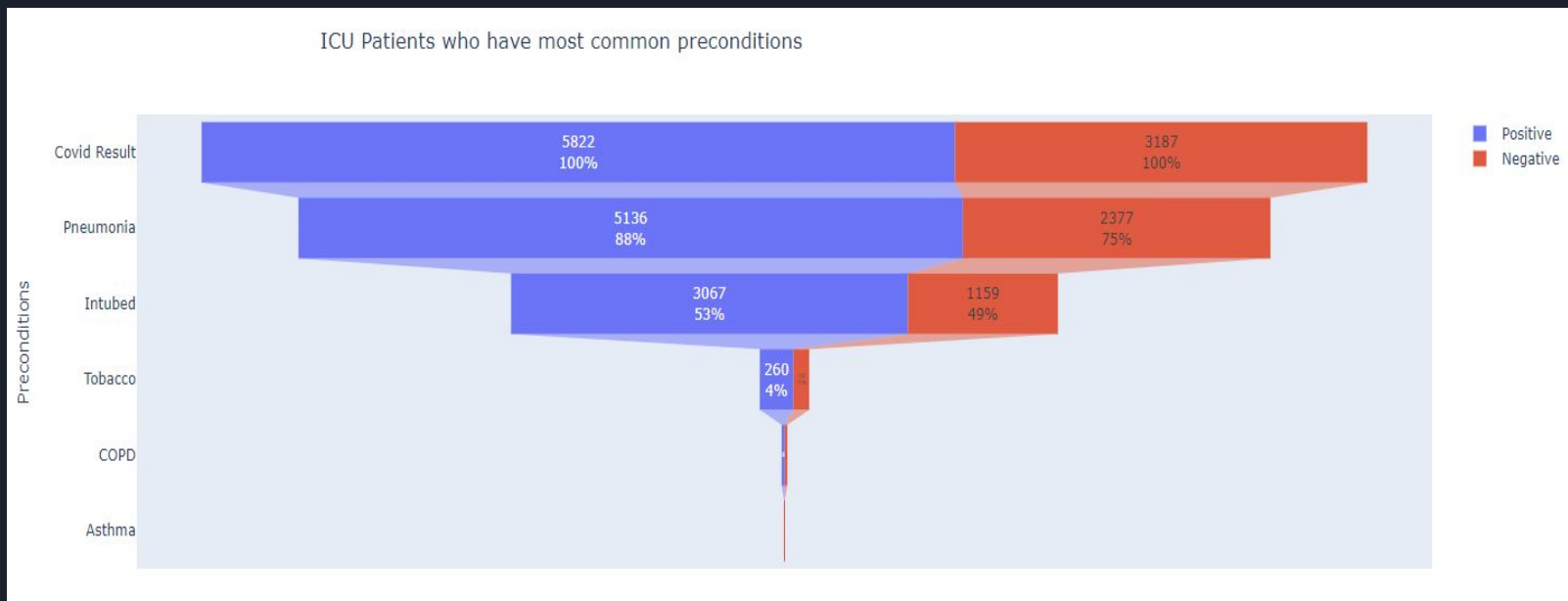
Result Analysis (2)

- If any patient is hospitalized with a positive test result, he should be given ICU urgently if they have pneumonia or they are infants.

- Large number of deaths are caused by COVID instead of other comorbidities.

- Mortality rate among patients who were hospitalized the day they started showing symptoms is high maybe due to high severity of Covid.

- The longer patients are in the hospital the mortality rate is reducing. It is likely that medical care given was working.

Thank you!