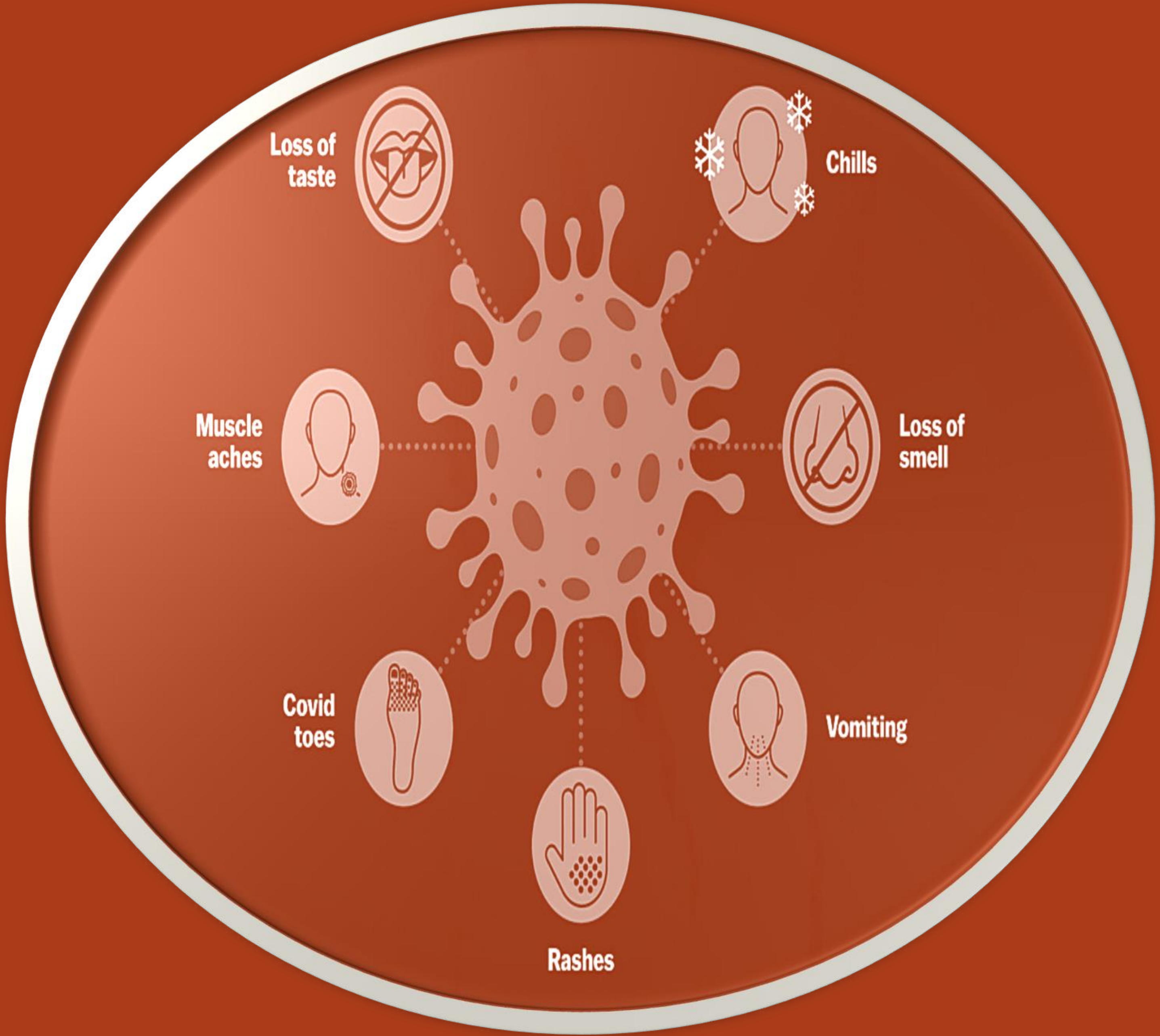
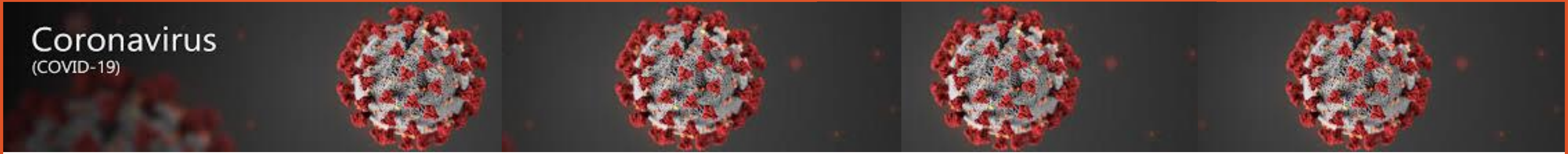


Covid 19 Diagnosis

11/06/2022



Agenda

- Problem Statement
- Project Approach
- Machine learning models
- Feature Importance
- Univariate Analysis
- Bivariate Analysis
- Insight
- Recommendation

Problem Statement

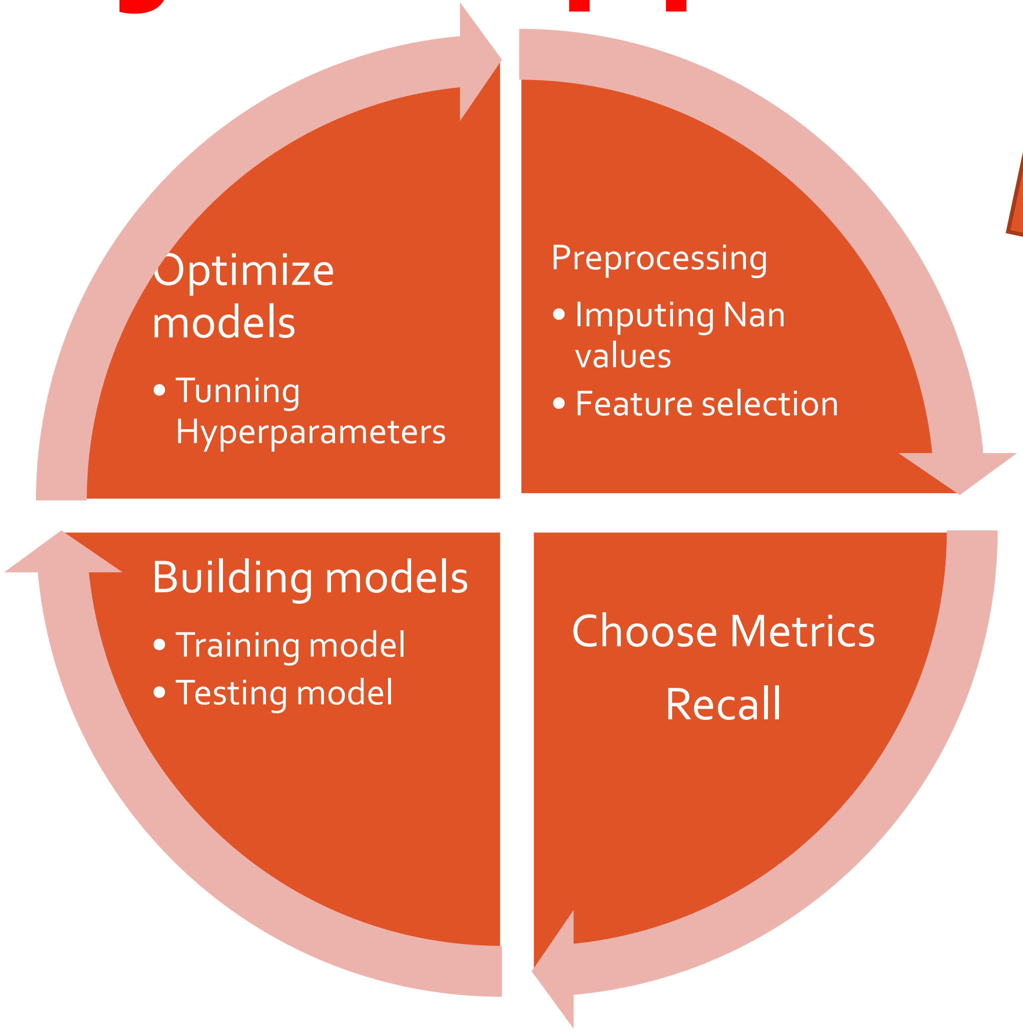
Motivation:

In pandemics and overwhelmed health system, the possible of limitation to perform tests to detect SARS-CoV-2 and test every case would be impractical. Tests results could be delayed even if small sample of population would be tested.

Objective:

Predicting the chances of being positive or negative for covid19 and identify the factors that influence it. Provide the recommendations to the hospital on how they can better manage the admission of patients to the general ward, semi-intensive unit, or intensive care unit.

Project Approach



Preprocessing

- Imputing Nan values
- Feature selection

Choose Metrics

Recall

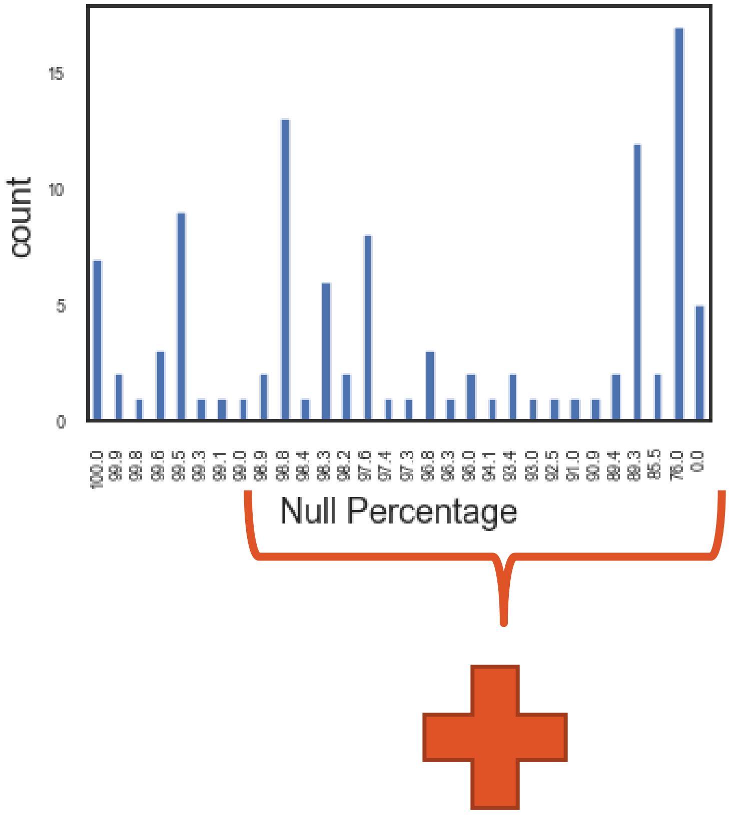
Building models

- Training model
- Testing model

Optimize models

- Tunning Hyperparameters

Preprocessing



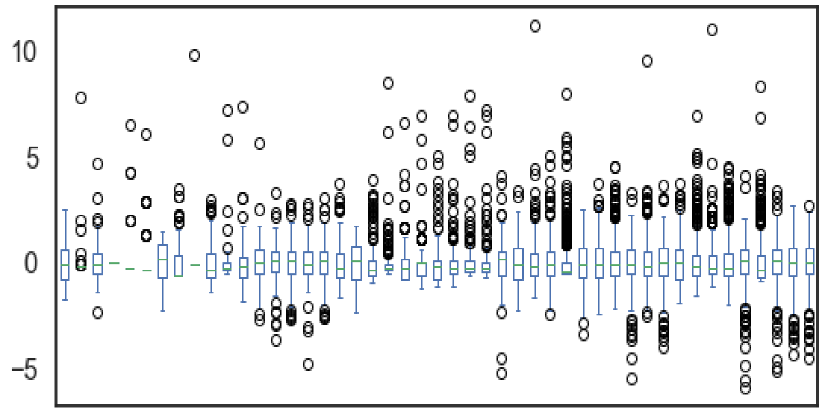
KnnImputer

Feature Selection:

- Drop features >99% nan values
- Drop low variance
- Drop high correlated features
- there are no outliers

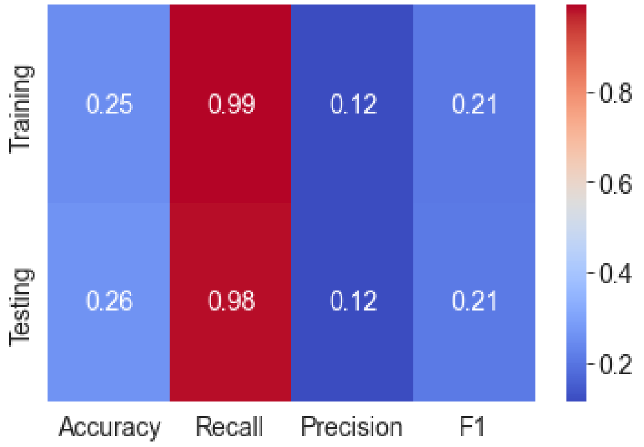
The final dataset:

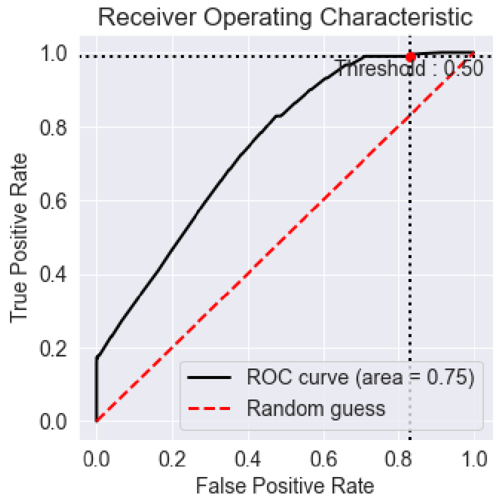
- Total row=5440
- Total columns=56
- Float features=51
- object features=5



Comparing the accepted Models

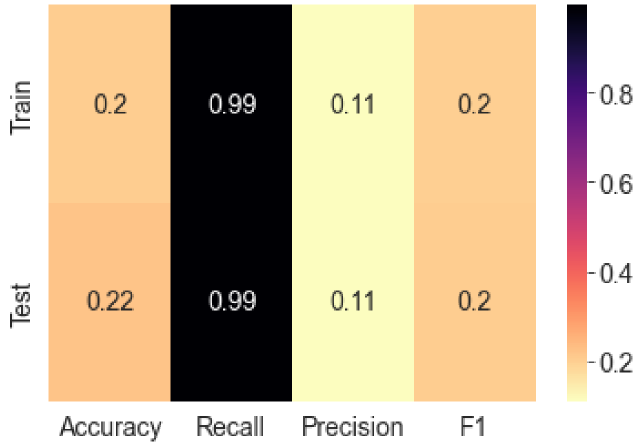
GradientBoost

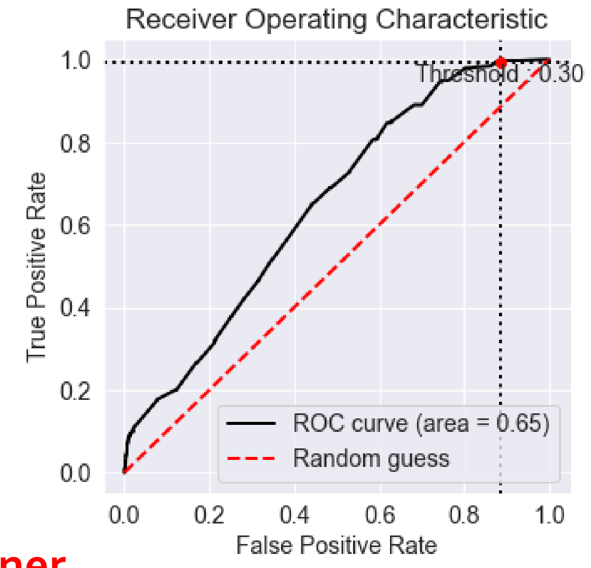




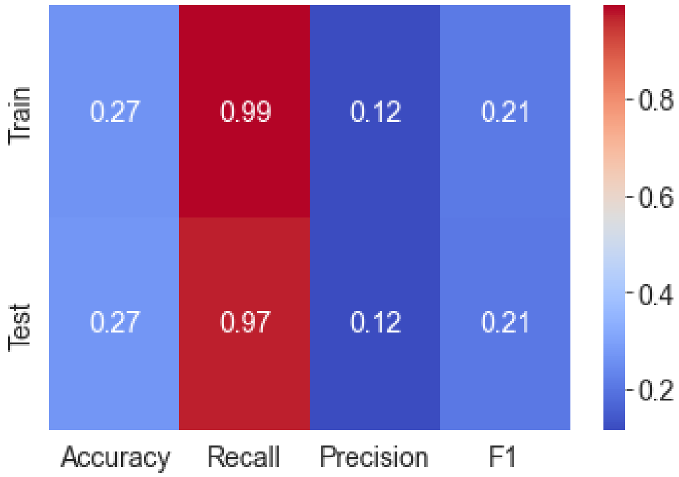
Winner

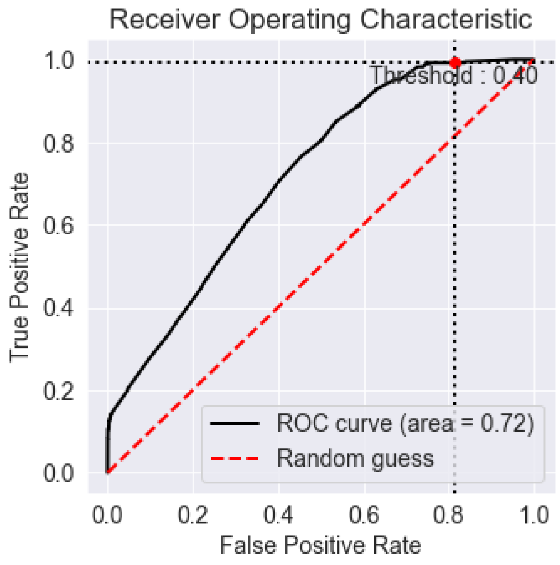
Logistic Regression



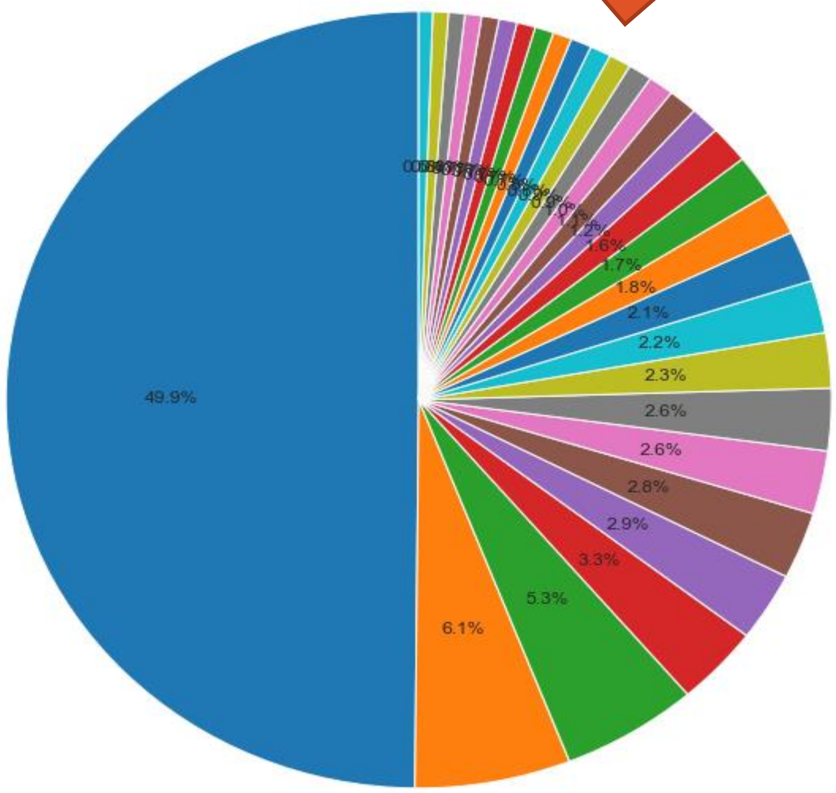


Random Forest

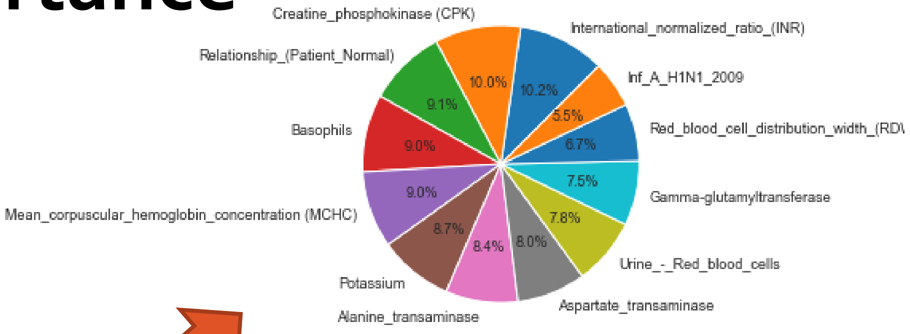




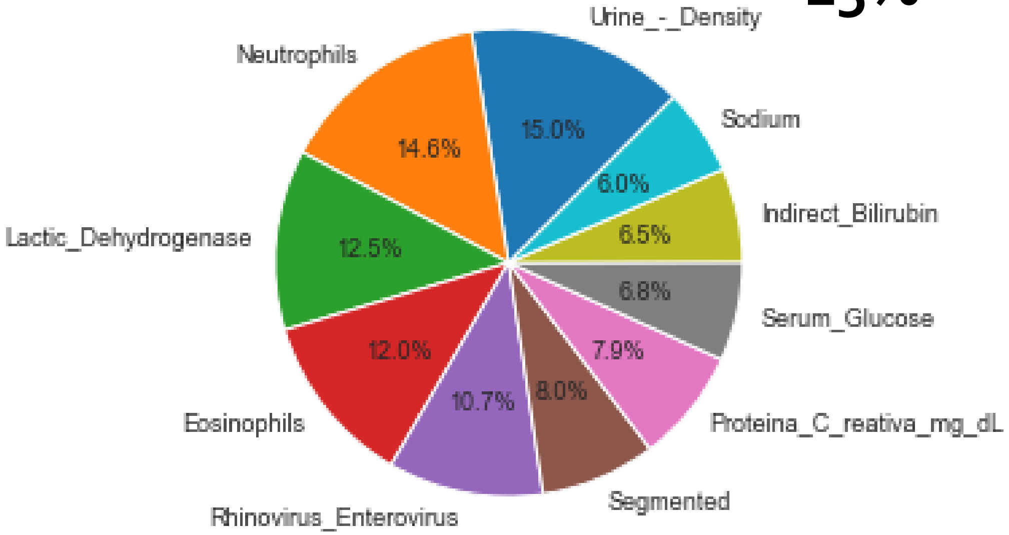
Feature importance



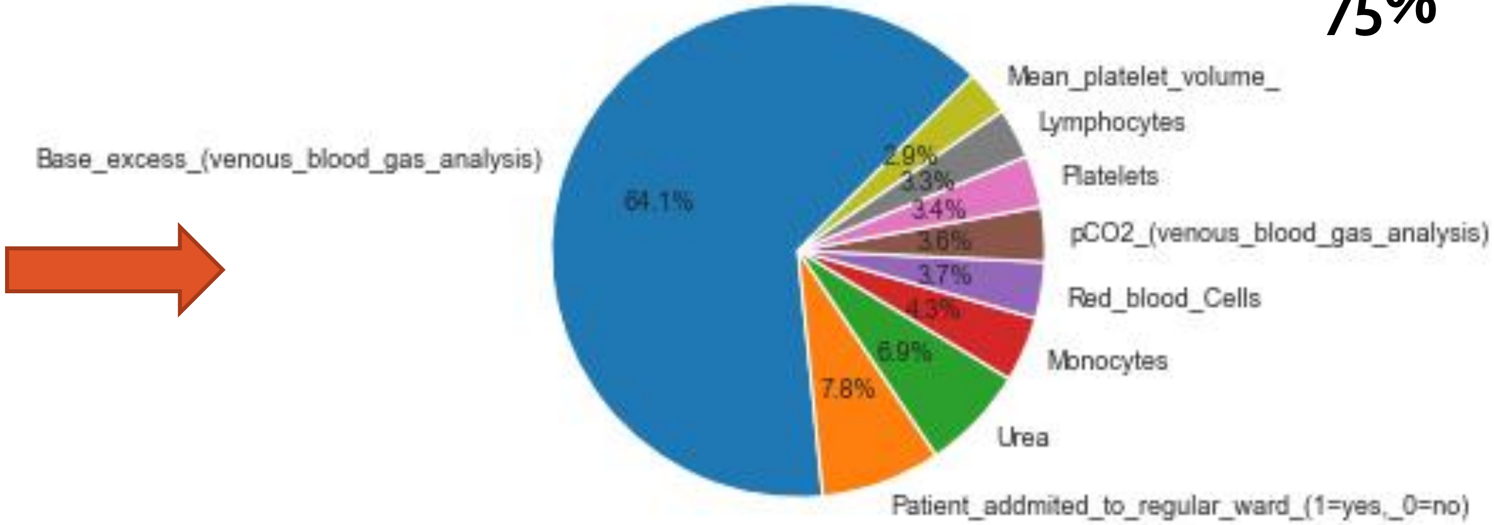
10%



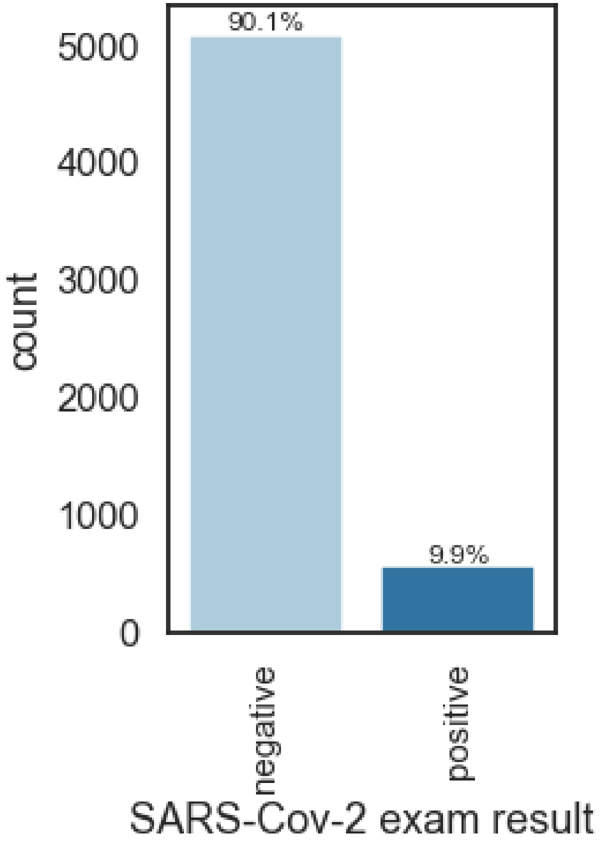
15%

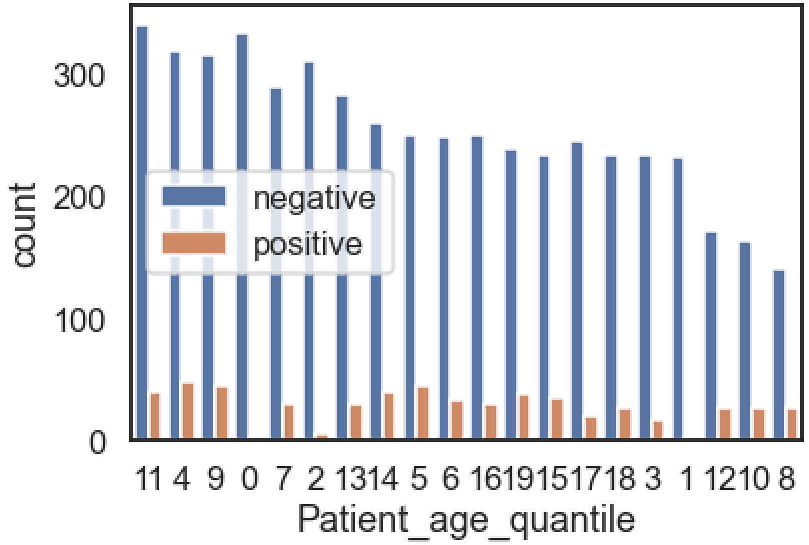


75%

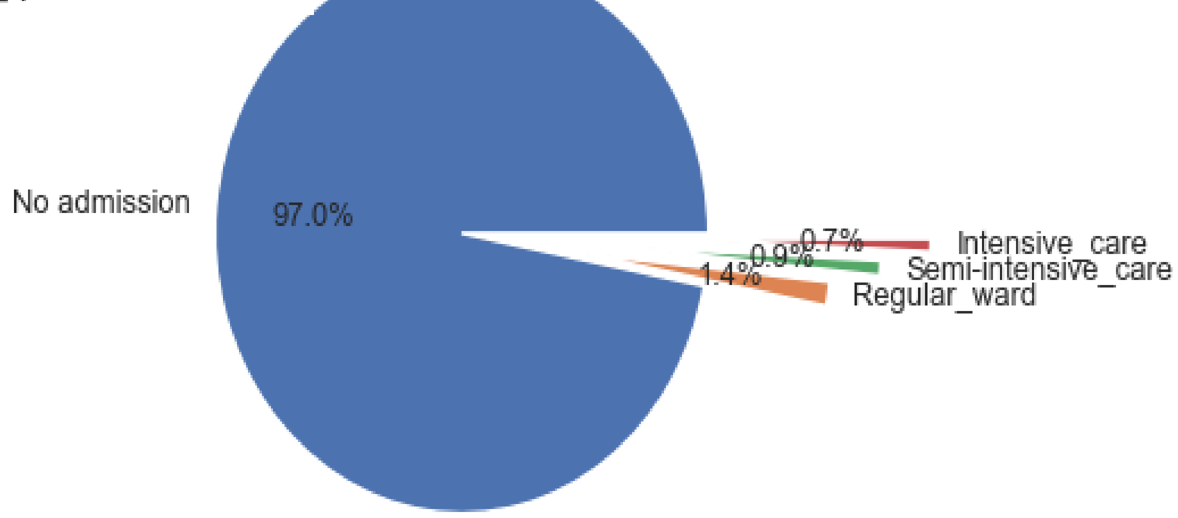


Univariate Analysis





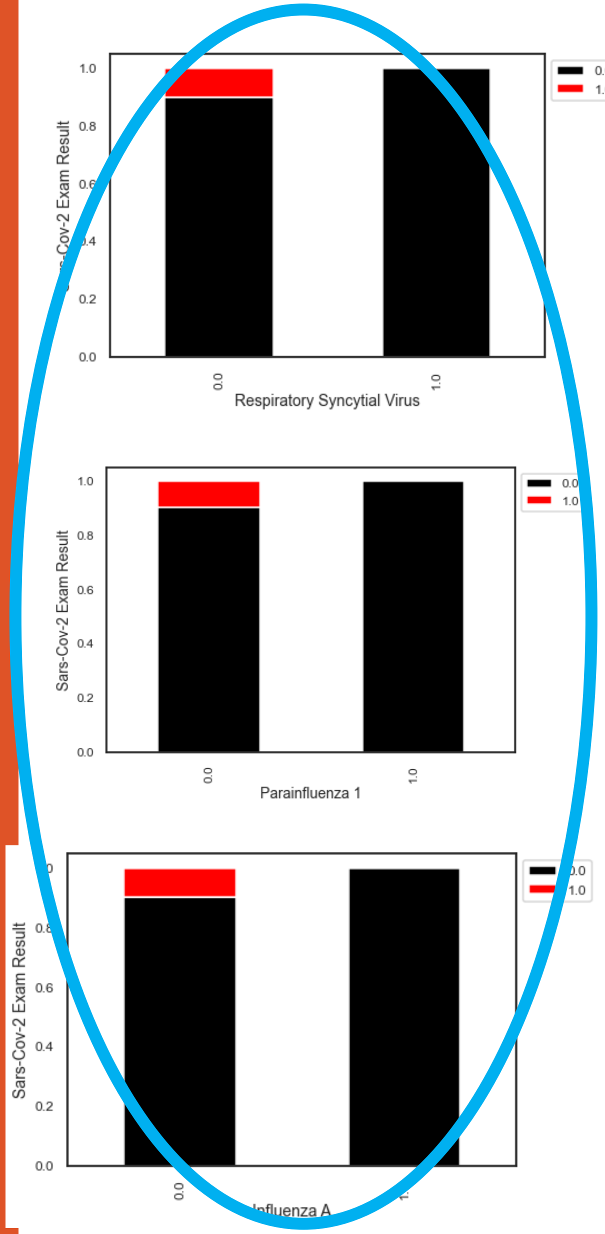
- Patients in 0,1 age quantile doesn't have covid.
- Patients in 11,4,and 9 quantiles have most of the covid cases

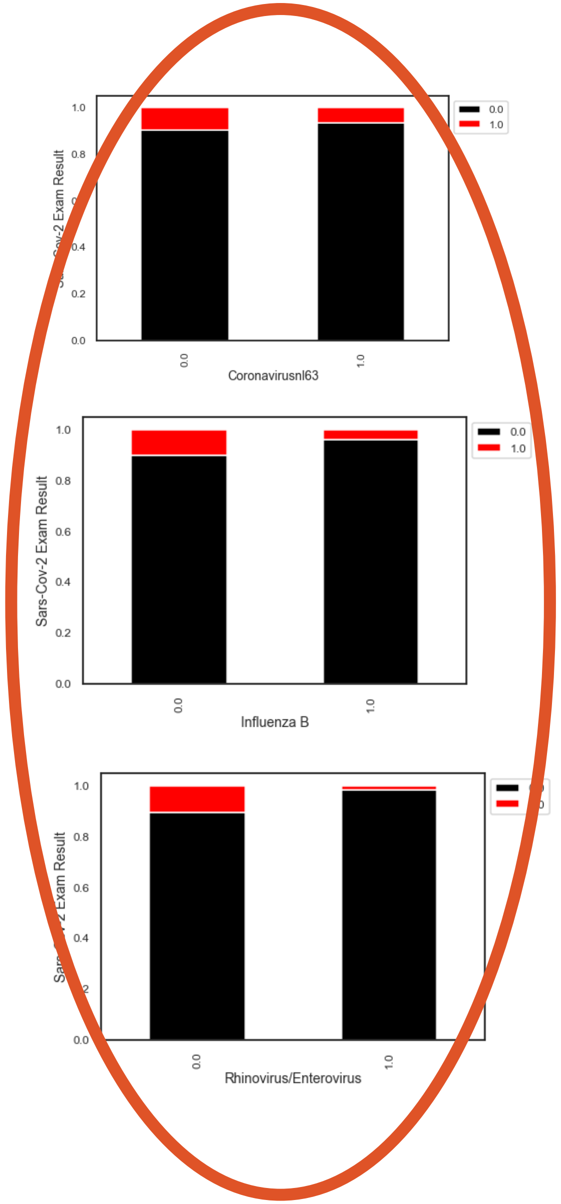


- Imbalanced data set with ratio 0.1
- 97% did not admitted in hospitals

Bivariate Analysis

Covid 19 and other viruses



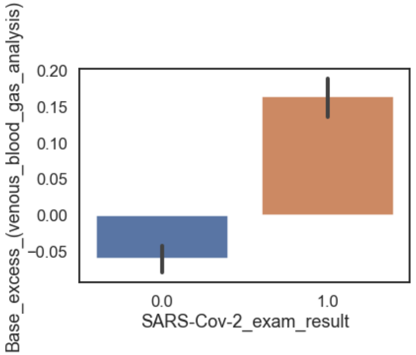


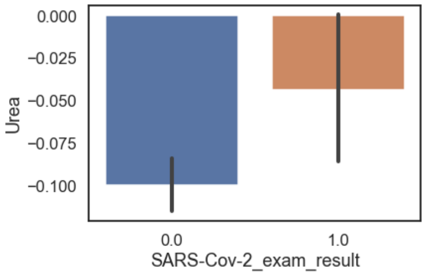
- when Influenza A, Parainfluenza 1, and Respiratory syncytial virus was not detected, negative, 10% of the cases tested positive to the covid 19
- Symptoms of Influenza A, Parainfluenza 1, and Respiratory syncytial virus most likely do not mix with Covid 19 symptoms.

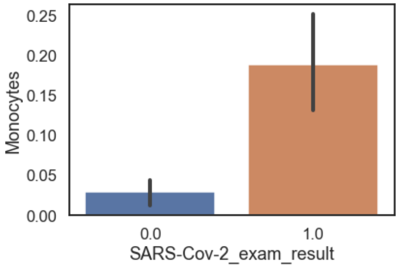
- Cases approved for Influenza B, Rhinovirus or Corona virrus63 have 2% ,1%,and 8% positive covid 19, respectively. While cases did not approve have 15% positive covid19.

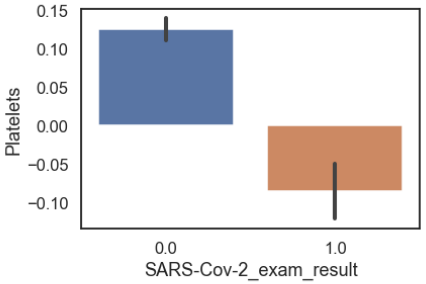
How the most important features relate to the Covid 19?

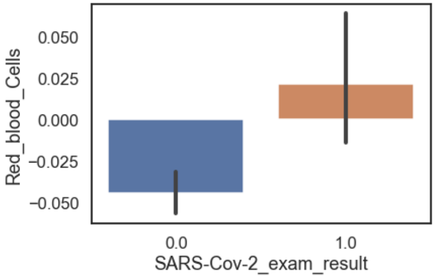
Bivariate Analysis

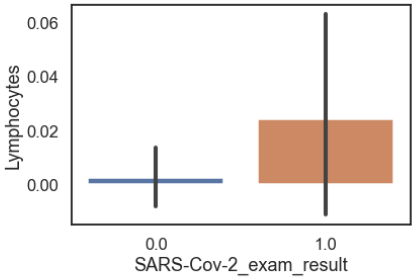


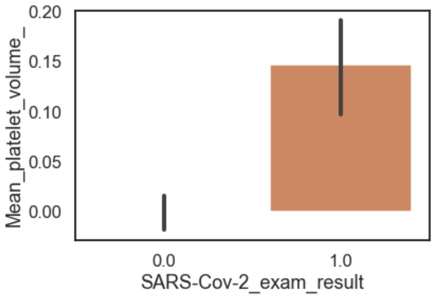


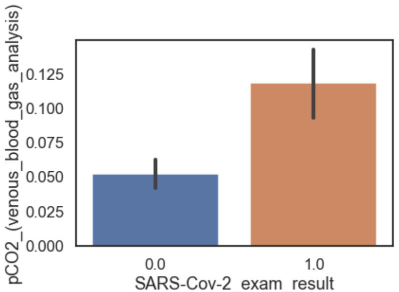


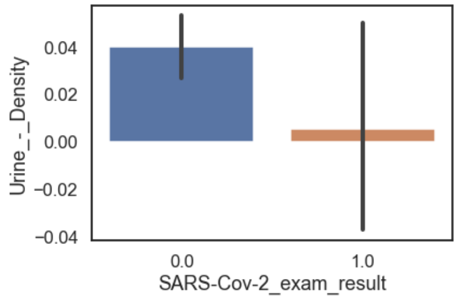




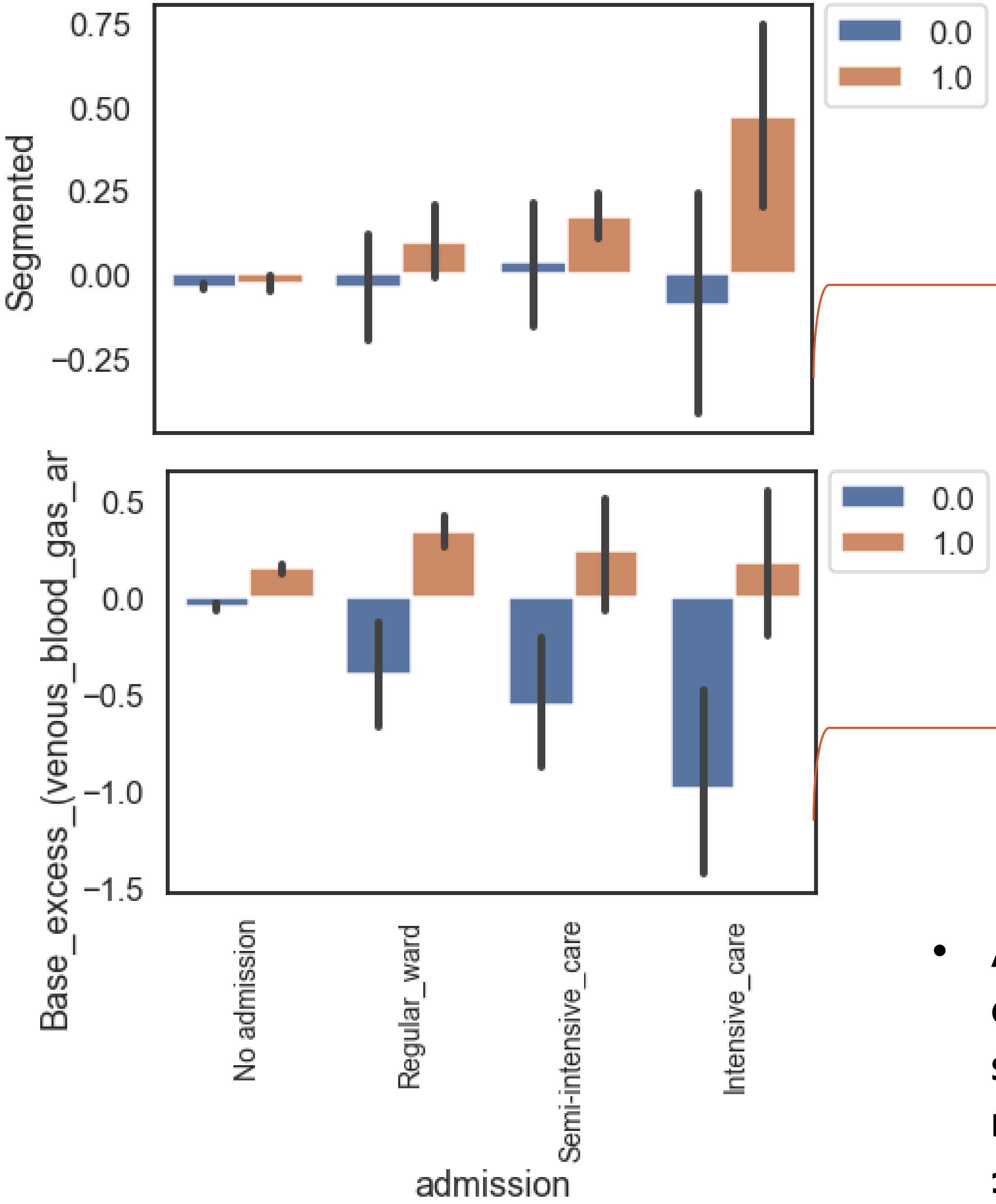






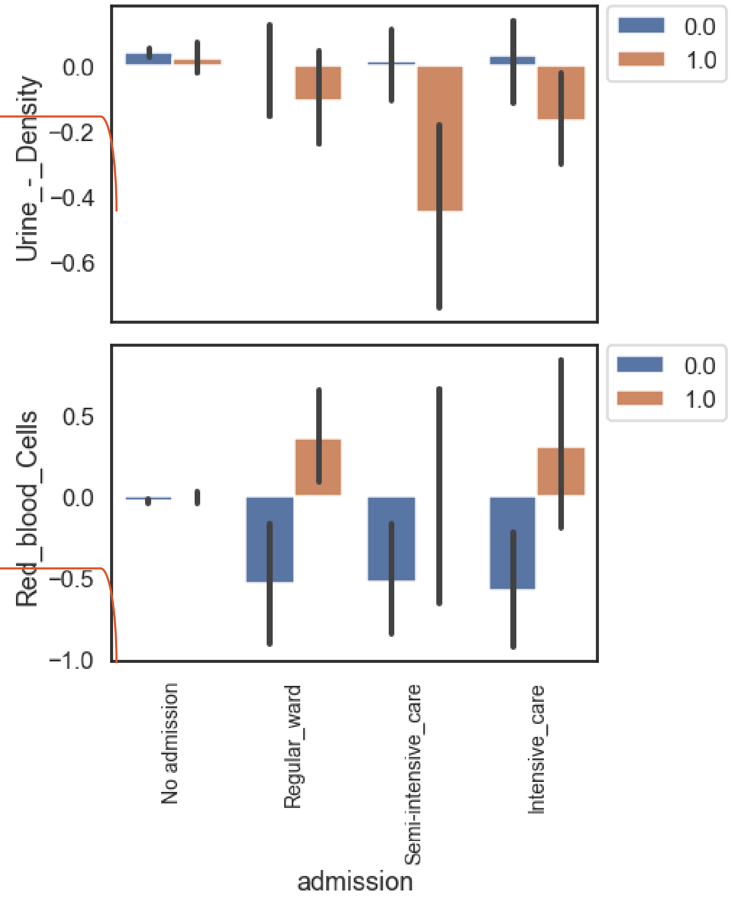


Bivariate Analysis



Covid related

Not Covid related



- Admission in hospitals could be related to the symptoms and test results more than covid 19 case.

Insights

- 97% of patients were not admitted in hospital, while 7% were admitted in intensive care.
- 45% of patients accepted to the Regular ward have a positive Covid19, while 18% of patients accepted to the semi-intensive care unit have a positive Covid 19, and 20% of patients accepted to the intensive care unit have a positive Covid 19.
- There are some other viruses could have similarity in symptoms with Covid19 such as Influenza B, Rhinovirus or Corona virus63.
- Patient age quantiles between 9 and 19 has higher positive covid19 cases than rest.
- Respiratory test are important factors in predicting the covid19. Base access (Vinous gas blood analysis) is a major variable in our study.

Recommendation

- Blood test is essential to track the infections and they are indicators of covid19. E.g., Platelets that shows values less than average for positive covid19, while for red_blood_cells test, the values were higher than average. We recommend investing in respiratory and blood tests for patients coming with symptoms because that's the key to track positive covid cases.
- Accepting in the hospital mostly is related to how extreme the lab test and how intense is the symptoms. Most patients admitted in regular ward has mild symptoms with only 45% tested positive covid19 test. Patients who accepted in intensive care mostly having other complications besides the covid 19 or just severe illness but not covid19.
- Additionally, these features should be more effectively investigated in further and future works.

Thank you!

Dataset

Data Report

file:covid19_dataset

File Format: xlsx

Size= row:5644, column:111

data types:float64:70; object: 37;int64: 4

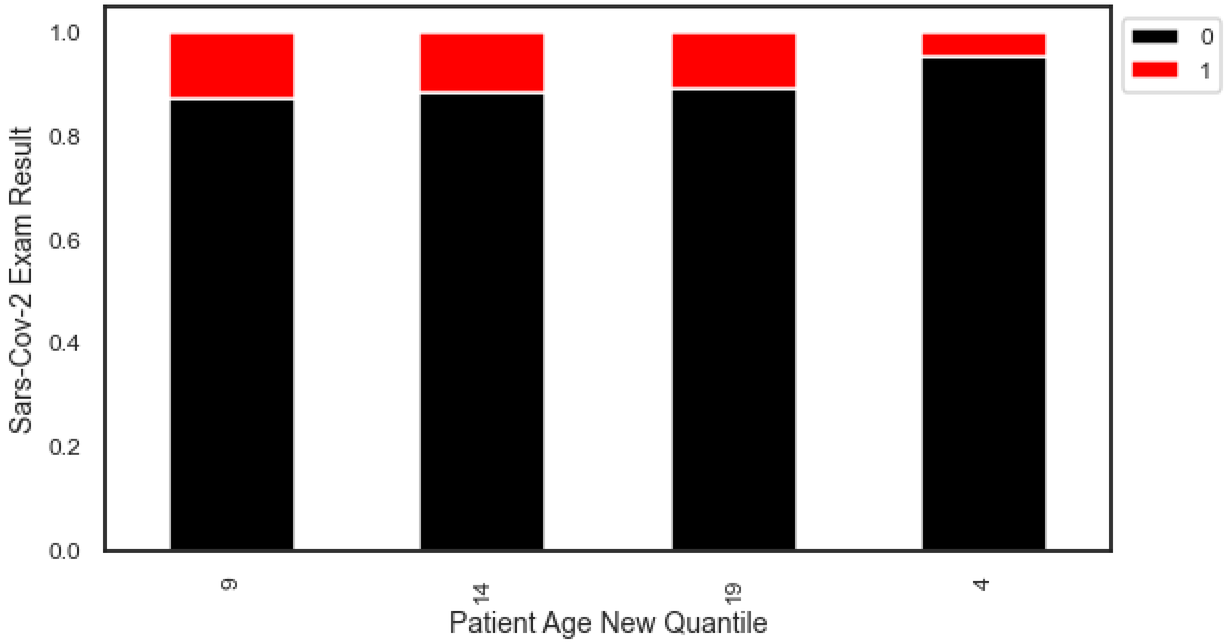
memory usage: 4.8+ MB

```
negative    5086
positive     558
Name: SARS-Cov-2 exam result, dtype: int64
NaN            4290
not_detected   1302
detected         52
Name: Respiratory Syncytial Virus, dtype: int64
NaN            4290
not_detected   1336
detected         18
Name: Influenza A, dtype: int64
NaN            4290
not_detected   1277
detected         77
Name: Influenza B, dtype: int64
NaN            4292
not_detected   1349
detected          3
Name: Parainfluenza 1, dtype: int64
NaN            4292
not_detected   1307
detected         45
Name: CoronavirusNL63, dtype: int64
NaN            4292
not_detected    973
detected        379
Name: Rhinovirus_Enterovirus, dtype: int64
NaN            4292
not_detected   1332
detected         20
Name: Coronavirus HKU1, dtype: int64z
```

```
NaN            4292
not_detected   1342
detected         10
Name: Parainfluenza 3, dtype: int64
NaN            4292
not_detected   1343
detected          9
Name: Chlamydophila pneumoniae, dtype: int64
NaN            4292
not_detected   1339
detected         13
Name: Adenovirus, dtype: int64
NaN            4292
not_detected   1333
detected         19
Name: Parainfluenza 4, dtype: int64
NaN            4292
not_detected   1343
detected          9
Name: Coronavirus229E, dtype: int64
NaN            4292
not_detected   1344
detected          8
Name: CoronavirusOC43, dtype: int64
NaN            4292
not_detected   1254
detected         98
Name: Inf A H1N1 2009, dtype: int64
NaN            4292
not_detected   1350
detected          2
```

Bivariate Analysis

Covid 19 variation with Age



Cases with age quantile between 9 and 4 has higher positive covid19 cases than all the other age quantile.
Ages less than 4 quantile has the lowest positive covid test.