Impact of Pre-existing Health Conditions and Demographic on COVID-19 Outcome

PRESENTED BY: ERIC MWANGI GITHINJI

Background

- Studies shows that pre-existing medical conditions can influence the severity of COVID-19. The study on “Impact of Pre-exisiting Health Conditions in Adults” found that, among COVID-19 cases, the three most common underlying health conditions are cardiovascular disease (32%), diabetes (30%), and chronic lung disease (18%). Among COVID-19 hospitalizations, the three most common underlying conditions are hypertension (57.7%), obesity (47.8%), and metabolic disease (42.9%).

> Ref: https://sph.uth.edu/research/centers/dell/legislative-initiatives/COVID-19%20Impact%20on%20pre-existing%20conditions_Adults_8_27_2021.pdf

- Demographic factors e.g. age and sex can as well influence the severity of COVID-19. For example, Adults are at a high risk of contracting the disease compared to younger people. Although adults aged older than 65 years represent only about 16% of the US population, they account for 31% of reported cases, 45% of hospitalizations, 53% of intensive care unit (ICU) admissions, and 80% of COVID-19–associated deaths.

> Ref: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9633621/#:~:text=Early%20and%20late%20reports%20showed,also%20differ%20in%20older%20adults.

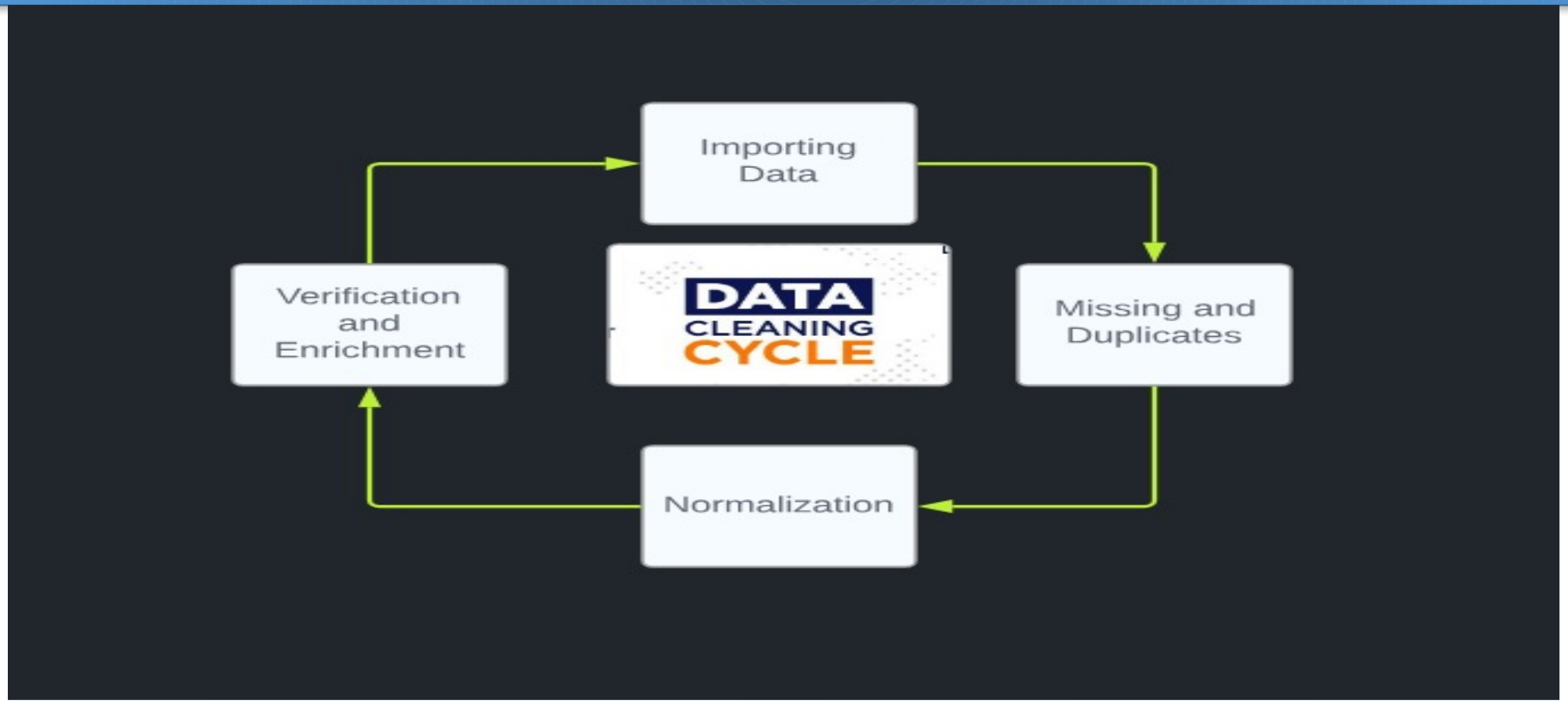
Analysis Objective

- To analyze the relationship between demographic factors (such as age and sex) and Covid-19 outcome.
- To investigate the impact of pre-existing medical conditions on Covid-19 outcomes.
- To study the timeline from symptom onset to death and its impact on patient outcomes.

Analysis Approach

- **Data Exploration and Cleaning**
  - Getting to know the dataset, structure, types of data in each column.
  - Identify and handle missing values, outliers and incorrect data entries.
- **Descriptive Analysis**
  - Summarizing data and finding patterns or relationships.
  - Use visualization to better understand the data.
- **Comparative Analysis**
  - Comparing different groups to identify significance differences.
- **Time-to-Event Analysis**
  - Survival analysis using cox proportional hazards model to analyze time-to-event.

Data Exploration and Cleaning



Covid-19 Mexico Data (10000 rows X 23 columns)

```
Rows: 10,000
Columns: 23
$ id                 <chr> "0dd249", "0957d9", "0ad2eb", "1646ba", "1e31dd", "16287e", "107c1f", "08cd…
$ sex                <chr> "Male", "Female", "Female", "Female", "Male", "Male", "Female", "Male", "Ma…
$ patient_type       <chr> "Outpatient", "Outpatient", "Outpatient", "Outpatient", "Outpatient", "Outp…
$ entry_date         <chr> "24-05-2020", "18-06-2020", "11-05-2020", "24-06-2020", "17-05-2020", "13-0…
$ date_symptoms      <chr> "18-05-2020", "16-06-2020", "11-05-2020", "19-06-2020", "16-05-2020", "07-0…
$ date_died          <chr> NA, NA, NA, NA, NA, NA, NA, NA, "15-04-2020", NA, "20-06-2020", NA, NA, NA,…
$ intubed            <chr> NA, NA, NA, NA, NA, NA, NA, NA, "Yes", NA, "No", NA, NA, NA, NA, NA, NA, NA…
$ pneumonia          <chr> "Yes", "No", "No", "No", "No", "No", "No", "No", "Yes", "No", "No", "No", "…
$ age                <dbl> 29, 45, 25, 11, 28, 9, 28, 58, 76, 30, 47, 34, 46, 38, 28, 68, 51, 44, 44, …
$ pregnancy          <chr> NA, "No", "No", "No", NA, NA, "No", NA, NA, "No", NA, NA, "No", "No", NA, N…
$ diabetes           <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "Yes", "No", "N…
$ copd               <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No…
$ asthma             <chr> "No", "No", "No", "Yes", "No", "No", "No", "No", "No", "No", "No", "No", "N…
$ inmsupr            <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No…
$ hypertension       <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "Ye…
$ other_disease      <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No…
$ cardiovascular     <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No…
$ obesity            <chr> "No", "No", "No", "No", "No", "No", "No", "No", "Yes", "Yes", "No", "No", "…
$ renal_chronic      <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No…
$ tobacco            <chr> "Yes", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "No", "N…
$ contact_other_covid <chr> "Yes", NA, "Yes", "Yes", "Yes", "Yes", "No", "Yes", "Yes", "Yes", NA, "No",…
$ covid_res          <chr> "Negative", "Negative", "Negative", "Positive", "Negative", "Negative", "Po…
$ icu                <chr> NA, NA, NA, NA, NA, NA, NA, NA, "No", NA, "No", NA, NA, NA, NA, NA, NA, NA,…
```

Summary Statistics

Summary Statistics for Numeric Variable

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
age	10000	43	20	-67	31	53	856

Numeric Data

Categorical Data

Summary Statistics for Categorical Variable

Variable	N	Percent
sex	10000	
... Female	4764	48%
... Femalee	111	1%
... Male	5073	51%
... Males	52	1%
patient_type	10000	
... Inpatient	2140	21%
... Outpatient	7860	79%
intubed	2137	
... No	1956	92%

Missing Values

Missing Values

```
List of 23
 $ id                 : int 0
 $ sex                : int 0
 $ patient_type       : int 0
 $ entry_date         : int 0
 $ date_symptoms      : int 0
 $ date_died          : int 9361
 $ intubed            : int 7863
 $ pneumonia          : int 1
 $ age                : int 0
 $ pregnancy          : int 5159
 $ diabetes           : int 35
 $ copd               : int 31
 $ asthma             : int 32
 $ inmsupr            : int 34
 $ hypertension       : int 32
 $ other_disease      : int 44
 $ cardiovascular     : int 34
 $ obesity            : int 31
 $ renal_chronic      : int 33
 $ tobacco            : int 34
 $ contact_other_covid: int 3107
 $ covid_res          : int 0
 $ icu                : int 7863
>
```

Duplicated Values

```
# A tibble: 8 × 23
```

	id	sex	patient_type	entry_date	date_symptoms	date_died	intubed	pneumonia	age	pregnancy	diabetes
	<chr>	<chr>	<chr>	<date>	<date>	<date>	<chr>	<chr>	<dbl>	<chr>	<chr>
1	0.00…	Fema…	Outpatient	2020-06-09	2020-06-05	NA	NA	No	34	No	No
2	0.00…	Male	Inpatient	2020-06-25	2020-06-22	NA	No	No	48	NA	No
3	0.00…	Fema…	Outpatient	2020-06-26	2020-06-15	NA	NA	No	41	No	No
4	0.00…	Male	Outpatient	2020-05-25	2020-05-23	NA	NA	Yes	16	NA	No
5	1.34…	Fema…	Inpatient	2020-06-13	2020-06-06	NA	No	Yes	55	No	No
6	0.00…	Fema…	Outpatient	2020-05-20	2020-05-18	NA	NA	No	40	No	No
7	1.34…	Fema…	Outpatient	2020-04-08	2020-04-06	NA	NA	No	47	No	No
8	0.00…	Male	Outpatient	2020-05-22	2020-05-16	NA	NA	No	40	NA	No

Variables Id, sex and entry_date merged to create a unique variable ID, hence eliminating duplicates.
Missing value handled differently for different variables.

Descriptive statistics GENDER*

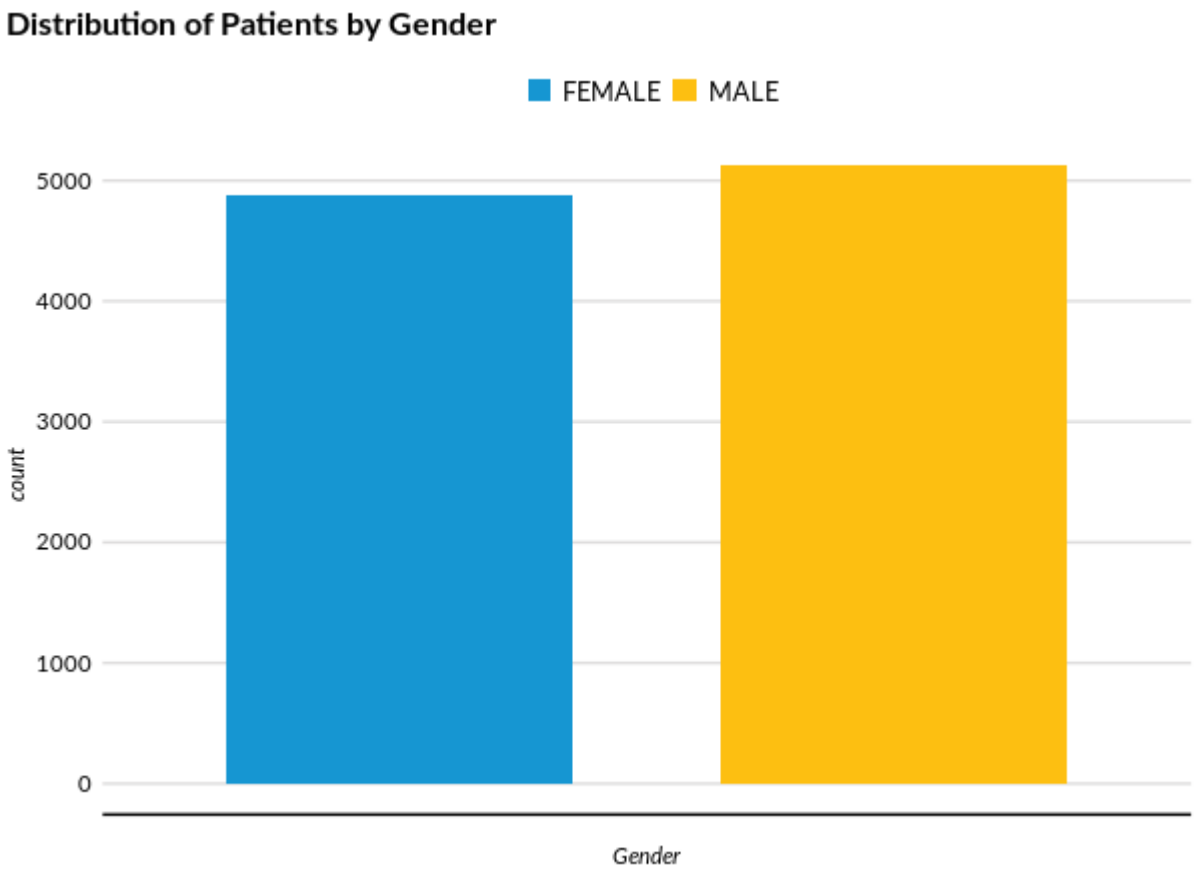


Figure 1: This bar chart represents the number of patients by Gender.

Male = 5125 (51.25%)
Female = 4875 (48.78%)
N= 10000

Descriptive statistics AGE*

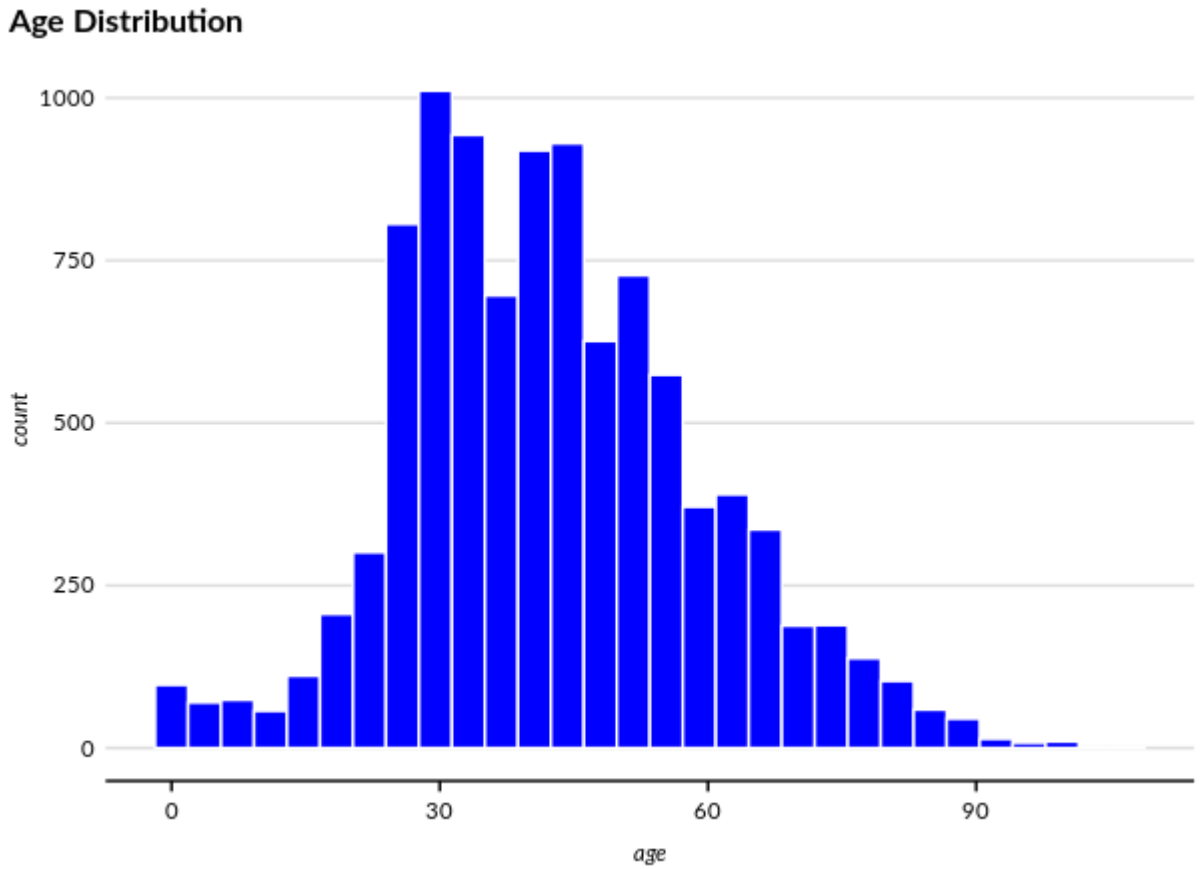


Figure 2: The histogram illustrates the age distribution of patients in the dataset

Mean=42.62
Sd = 16.82
Min= 0
Max= 107
N= 10000

Descriptive statistics Symptoms Period

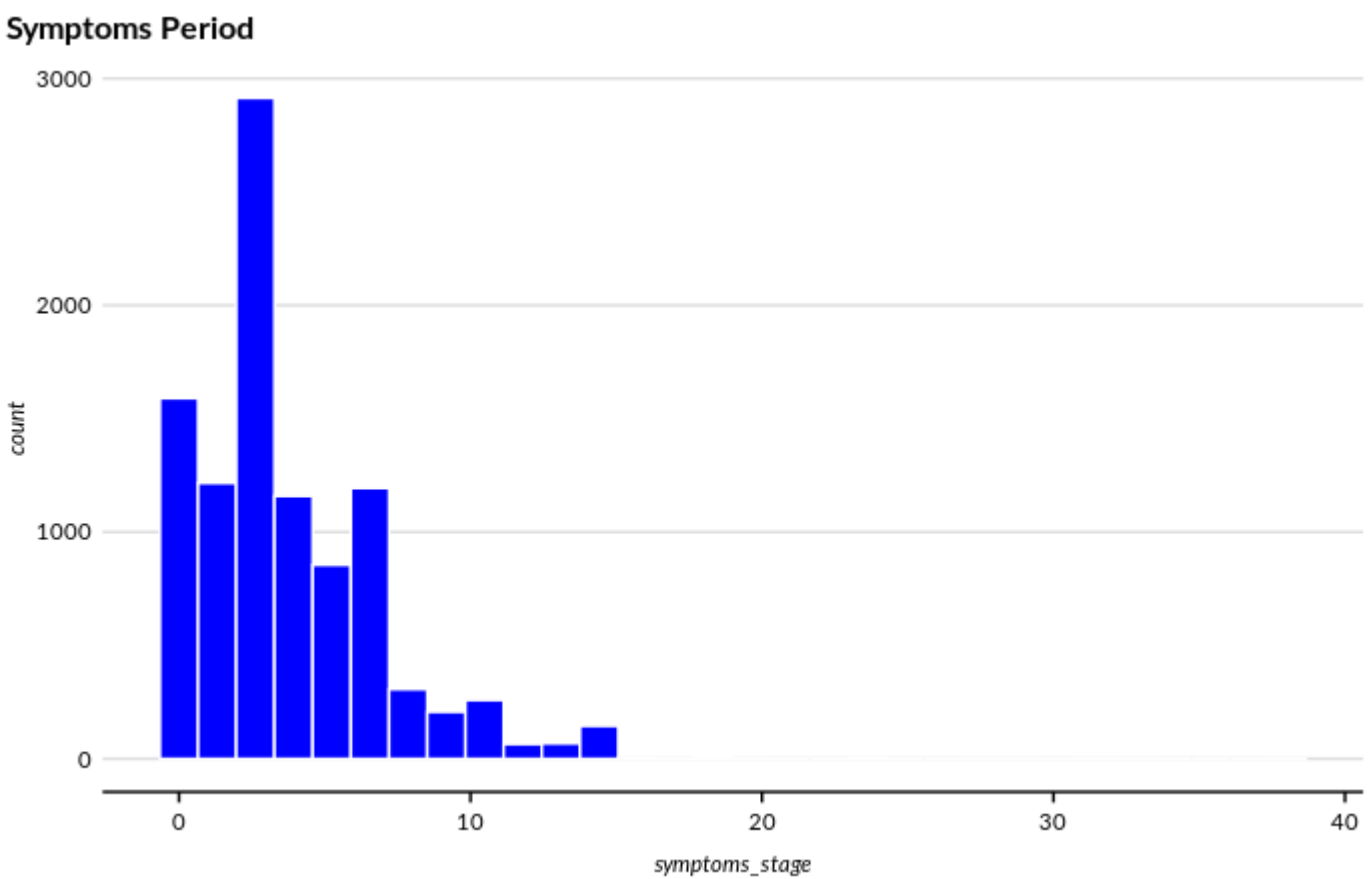


Figure 3: The histogram illustrates days before a patient vist a unit care after symptopsm

Mean=3.64
Sd = 3.24
Min= 0
Max= 38
N= 10000

Days between symptoms and entry date.

Descriptive statistics
Survival Period

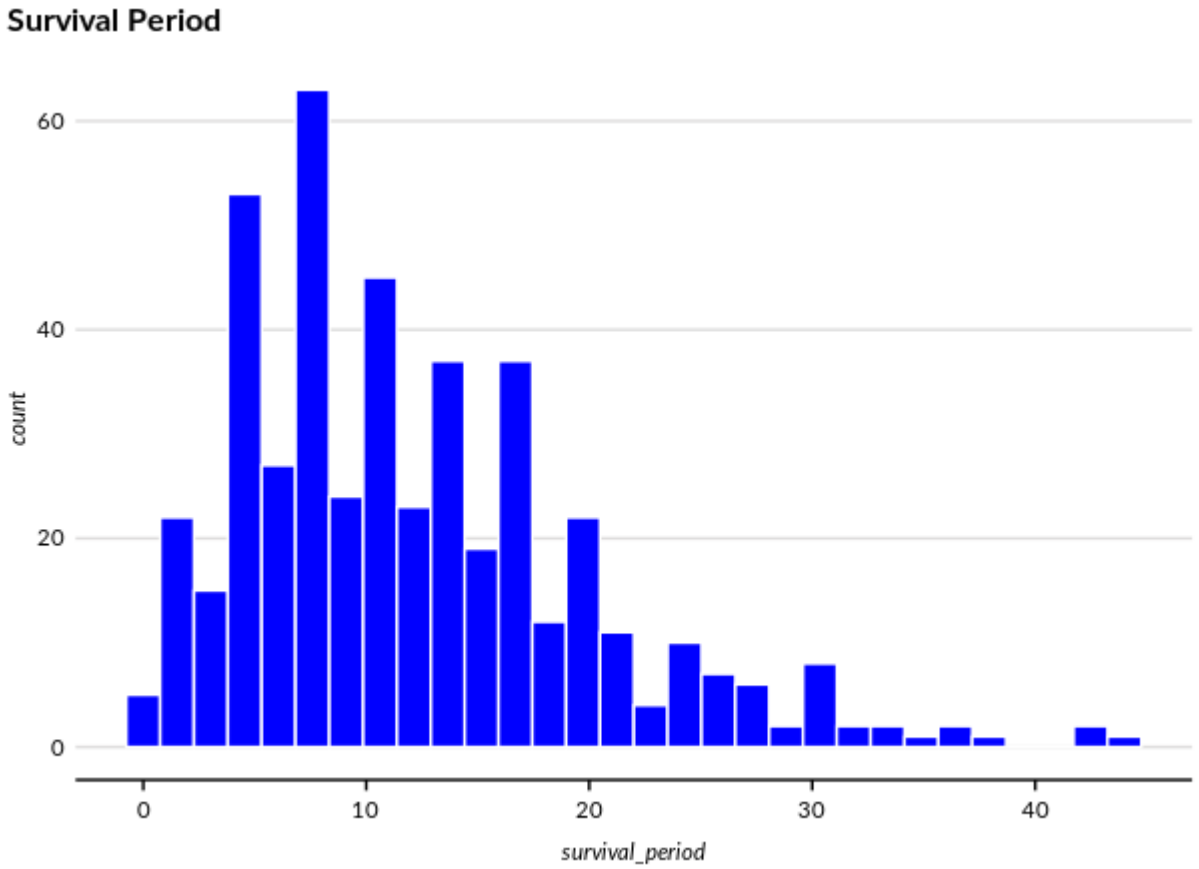


Figure 4: The histogram illustrates the distribution of survival period

Mean=12.06
Sd = 7.89
Min= 0
Max= 44
N= 463

Days between symptoms and death of a patient.

Descriptive statistics COVID-19 Results

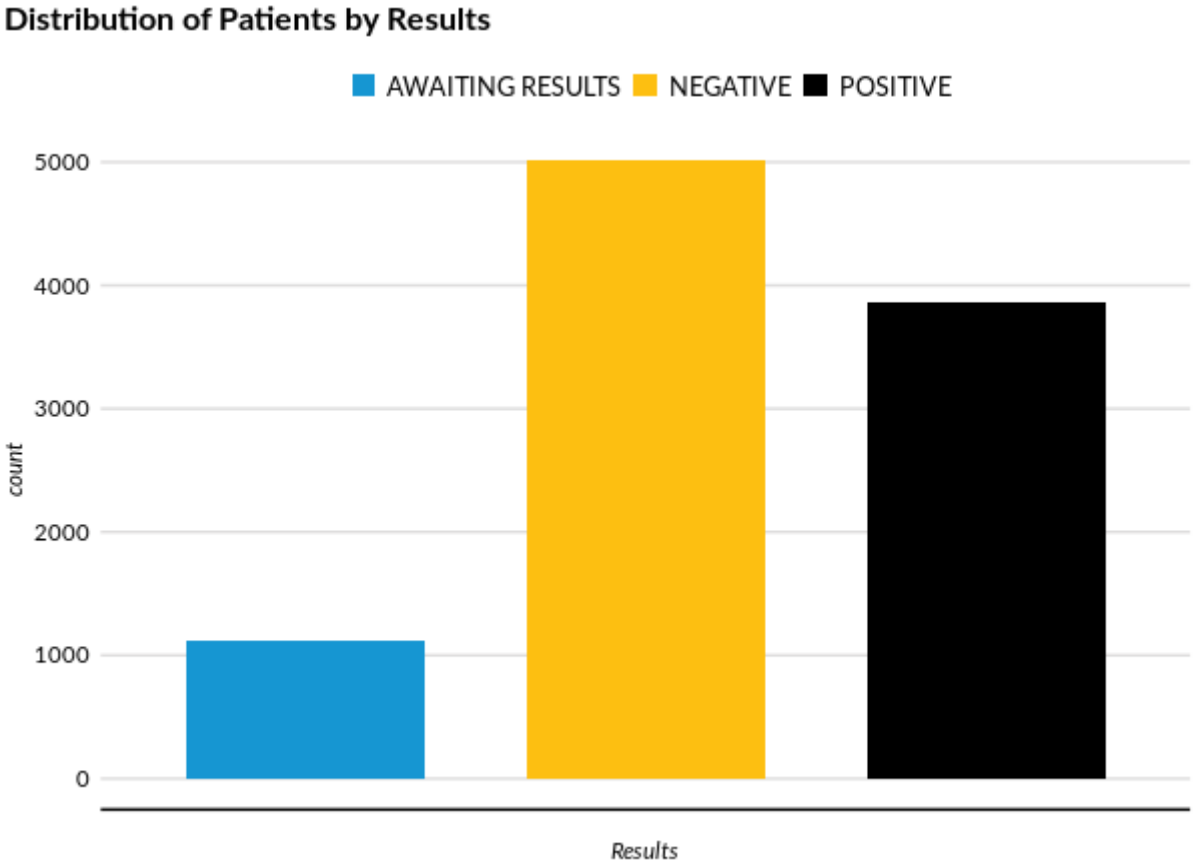


Figure 6: This bar chart represents the number of patients by their test results.

Awaiting Results=1122 (11.22%)
Negative = 5013 (50.13%)
Positive= 3865 (38.65%)
N= 10000

Correlation Analysis

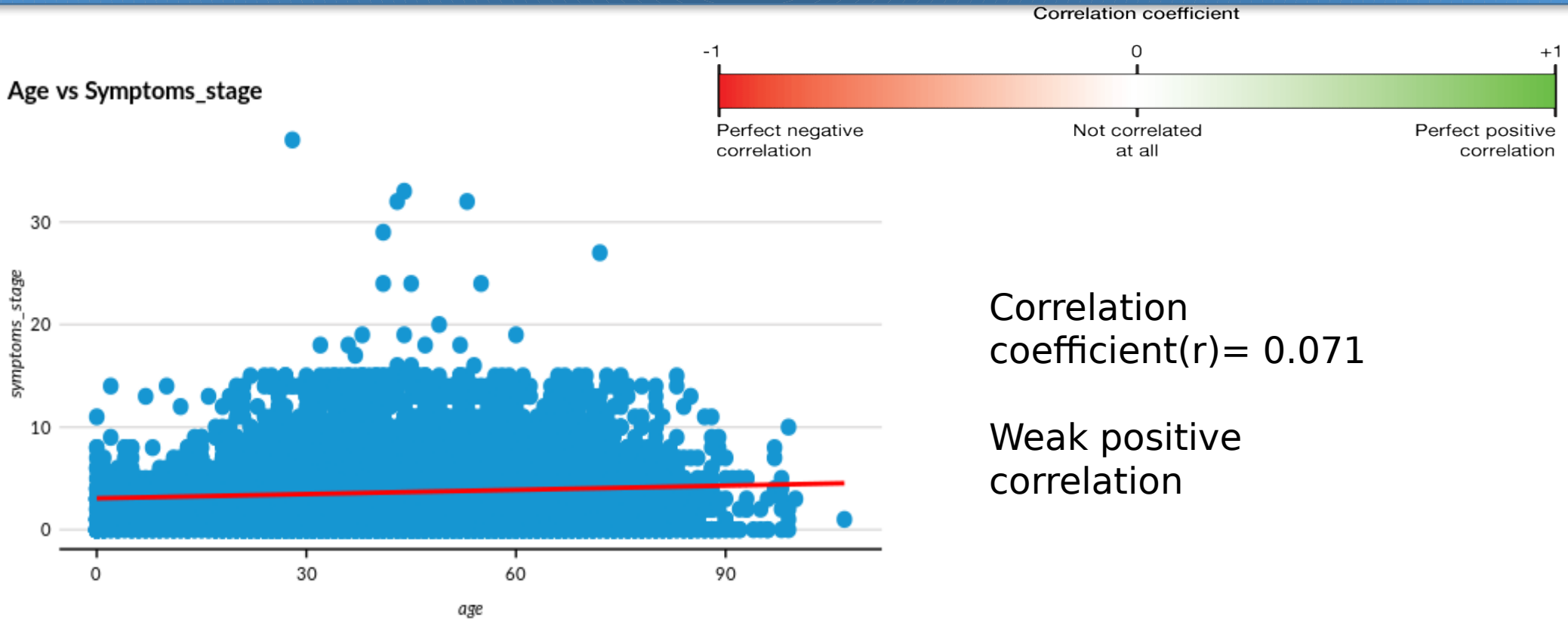


Figure: The week plot show a positive correlation between age and sympton_stage

Correlation coefficient(r)= 0.071

Weak positive correlation

Comparative Analysis
Group Means

Difference in Mean (Age)

Covid Results	AGE
POSITIVE	45.58
AWAITING RESULTS	43.85
NEGATIVE	40.05

Difference in Mean (Symptoms_stage)

Covid Results	Symptoms_stage
POSITIVE	4.21
AWAITING RESULTS	3.98
NEGATIVE	3.12

Is the difference significant?

Comparative Analysis
One-Way ANOVA

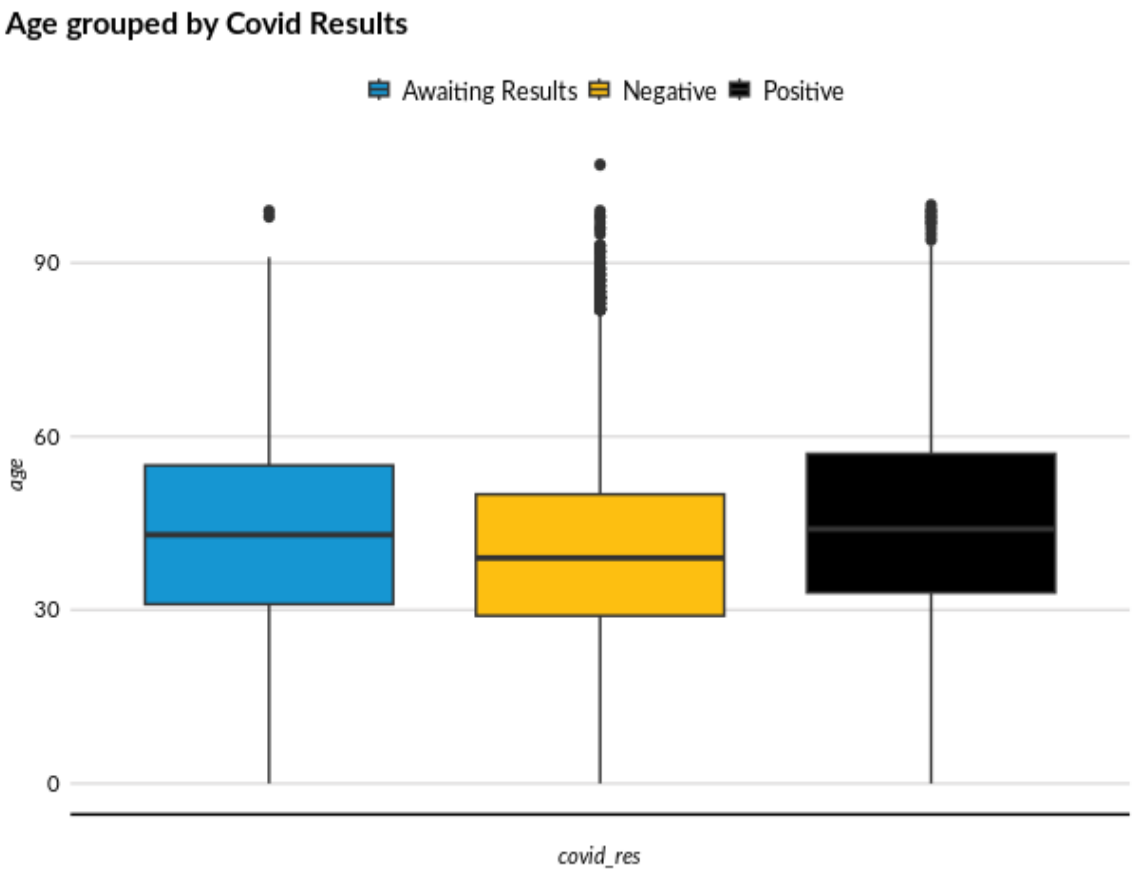


Figure 7: The box plot shows a slight differemce in age among the three groups

Anova assumptions: Homogeneity of Variance (Equal variance between groups)
Bartlett's test: H0> All groups have equal variance
H1> Not all groups have equal variance
P-value =0.4418 > 0.05. Fail to reject null hypothesis.

Model Summary

Parameter	Sum_Squares	df	Mean_Square	F	p
covid_res	68759.24	2	34379.62	124.54	< .001
Residuals	2.76e+06	9997	276.06		

Anova Table (Type 1 tests)

Tukey multiple comparisons of means

	diff	lwr	upr	p adj
Negative-Awaiting Results	-3.805	-5.091	-2.519	0.000
Positive-Awaiting Results	1.728	0.407	3.049	0.006
Positive-Negative	5.533	4.700	6.367	0.000

Comparative Analysis
Welch’s ANOVA

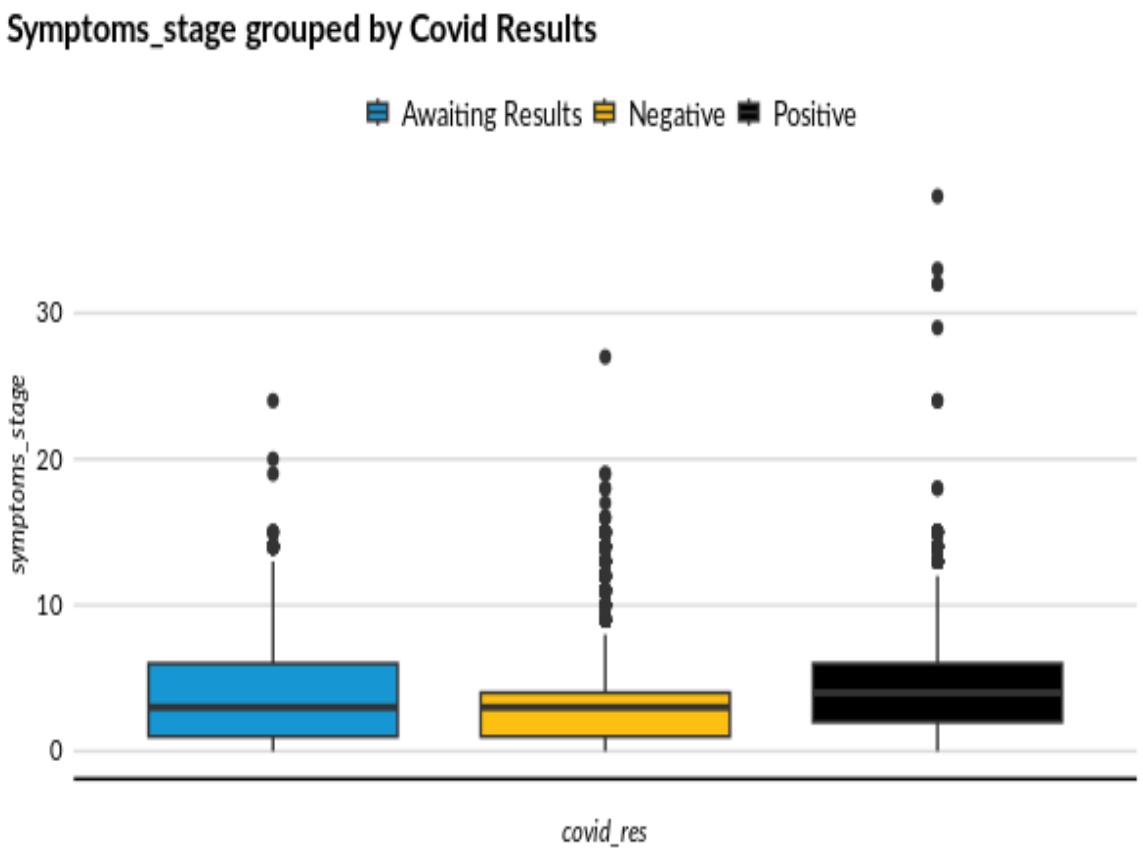


Figure 8: The box plot shows a slight differemce in symptoms_stage among the three groups

Anova assumptions: Homogeneity of Variance (Equal variance between groups)
Bartlett’s test: H0> All groups have equal variance
H1> Not all groups have equal variance
P-value =0.00 > 0.05. Reject null hypothesis.

Welch's ANOVA

F	df	df_error	p	Method
132.932	2	3035.957	0	One-way analysis of means (not assuming equal variances)

Post-Hoc Welch's Anova

.y.	group1	group2	n1	n2	statistic	df	p	p.adj
symptoms_stage	AWAITING RESULTS	NEGATIVE	1122	5013	7.83	1528.73	0.00	0.00
symptoms_stage	AWAITING RESULTS	POSITIVE	1122	3865	-2.00	1833.90	0.05	0.14
symptoms_stage	NEGATIVE	POSITIVE	5013	3865	-15.73	7666.32	0.00	0.00

Comparative Analysis Proportion Tables

Diabetic Patients

Covid results for patiens with diabitics

COVID-19 Results	Condition: No	Condition: Yes
AWAITING RESULTS	0.1114923	0.1163934
NEGATIVE	0.5188107	0.3762295
POSITIVE	0.3696970	0.5073770
Total	1.0000000	1.0000000

YES

- Approximately 11.64% are still awaiting results.
- About 37.62% have tested negative for Covid-19.
- Around 50.74% have tested positive for Covid-19

Pneumonia Patients

Covid results of patiens with pnemonia

COVID-19 Results	Condition: No	Condition: Yes
AWAITING RESULTS	0.1159489	0.0917906
NEGATIVE	0.5356129	0.3135100
POSITIVE	0.3484382	0.5946994
Total	1.0000000	1.0000000

YES

- Approximately 9.17% are still awaiting results.
- About 31.35% have tested negative for Covid-19.
- Around 59.46% have tested positive for Covid-19

Time-to-Event Analysis (Survival Analysis)

Significant Predictors in Survival Model

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
sexMALE	0.2352	1.2652	0.0856	2.7493	0.0060
patient_typeOutpatient	-2.4677	0.0848	0.1572	-15.6954	0.0000
intubedYes	0.7405	2.0969	0.1118	6.6254	0.0000
pneumoniaYes	0.6627	1.9400	0.1048	6.3227	0.0000
age	0.0308	1.0313	0.0028	11.1274	0.0000
other_diseaseYes	0.3115	1.3654	0.1523	2.0452	0.0408
renal_chronicYes	0.3309	1.3922	0.1518	2.1804	0.0292
covid_resNEGATIVE	-0.4287	0.6514	0.1819	-2.3567	0.0184
covid_resPOSITIVE	0.4706	1.6010	0.1608	2.9274	0.0034
symptoms_stage	-0.0298	0.9706	0.0118	-2.5268	0.0115

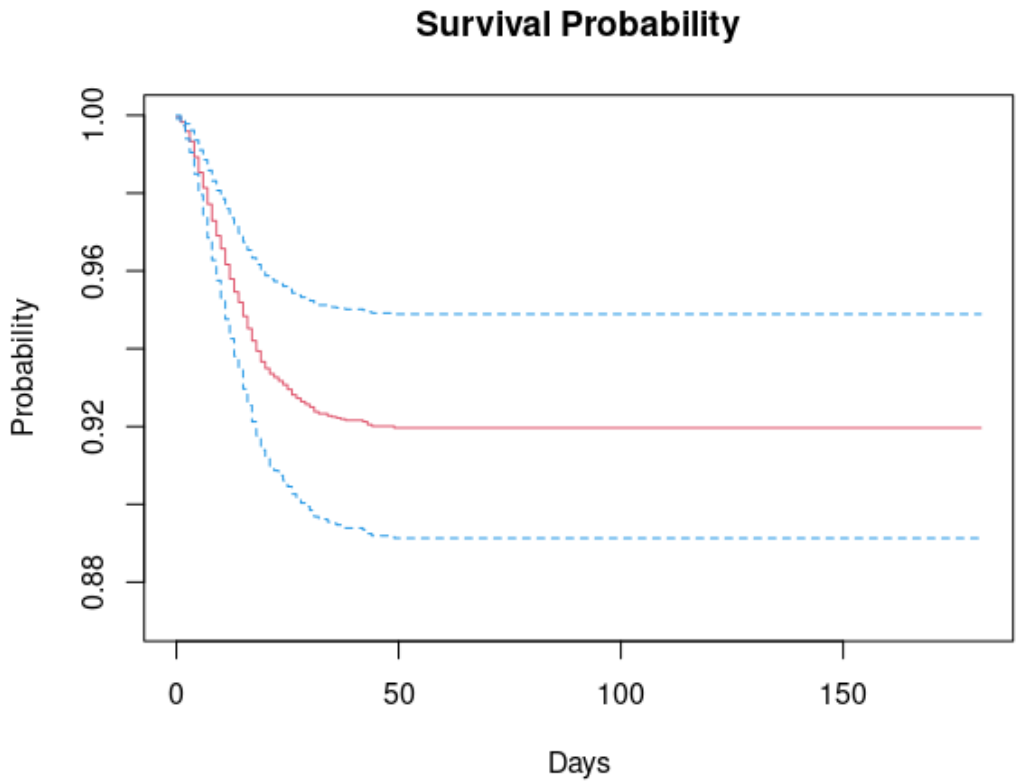
A positive coefficient indicates an increase in the hazard rate, and a negative coefficient indicates a decrease in the hazard rate.

At a given point in time someone who is male is 1.2652 times likely to die as someone who is female adjusting for other variable.

At a given point in time someone who is male is 26.52% more times likely to die as someone who is female adjusting for other variable.

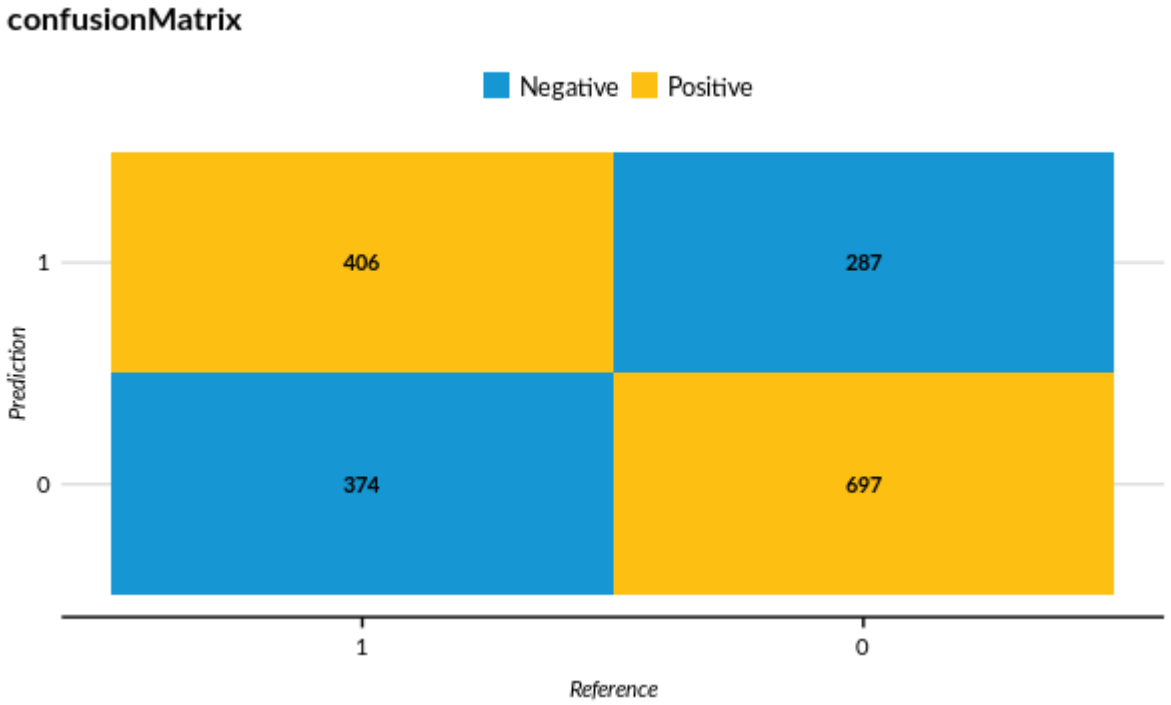
At a given point in time the probability of dying for someone who is 1 year 0ld is 1.0313 times higher than someone who is 1 year young.

Time-to-Event Analysis (Survival Analysis)



The y-axis represents the probability of survival. A survival probability of 1 means all individuals are alive, while a survival probability of 0 means all individuals have experienced the event

Classification Model




```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 697 374
         1 287 406

               Accuracy : 0.6253
                 95% CI : (0.6022, 0.6479)
    No Information Rate : 0.5578
    P-Value [Acc > NIR] : 5.358e-09

                  Kappa : 0.2315

 Mcnemar's Test P-Value : 0.0008228

            Sensitivity : 0.7083
            Specificity : 0.5205
         Pos Pred Value : 0.6508
         Neg Pred Value : 0.5859
             Prevalence : 0.5578
         Detection Rate : 0.3951
   Detection Prevalence : 0.6071
      Balanced Accuracy : 0.6144

       'Positive' Class : 0
```

Summary

Demographic Factors

- The average age of patients diagnosed with COVID-19 is notably higher, indicating a potential age-related vulnerability to the virus.

Pre-existing Conditions

- Our data suggests a correlation between pre-existing conditions such as diabetes and pneumonia, and an increased number of positive COVID-19 cases. This means that these conditions may make people more likely to get sick from the virus..

Survival Probability

- The presence of pre-existing conditions appears to negatively impact the survival rate among COVID-19 patients, suggesting these conditions could complicate the disease progression.
- A delay in seeking medical attention seems to decrease the survival probability for COVID-19 patients. This underscores the importance of timely medical intervention in improving patient outcomes.

Recommendations

- Public health measures should consider age and pre-existing conditions as significant factors of COVID-19 outcome. This could include prioritizing older individuals and those with pre-existing for vaccination and implementing targeted protective measures for such cases.
- Healthcare providers should be prepared to provide enhanced care for patients with Pre-existing conditions. This could include closer monitoring, more aggressive treatment strategies, and comprehensive post-recovery care.
- Public health campaigns should emphasize the importance of seeking medical attention at the earliest signs of symptoms, and healthcare systems should be prepared to respond swiftly to new cases.


THANK YOU