Patient Vital Signs Analysis for Risk Level Classification

Problem Statement

Healthcare professionals rely on vital signs to assess patient conditions and prioritize care. This dataset contains vital sign measurements (such as heart rate, respiratory rate, oxygen saturation, blood pressure, and temperature) along with consciousness level and oxygen support status for individual patients. Each patient is assigned a risk level — Low, Medium, or High — based on their clinical condition.

The goal of this analysis is to explore patterns and correlations within the patient data that can help identify key indicators associated with high-risk patients. By analyzing this dataset using Python and Pandas, we aim to:

Understand the distribution and relationship between different vital signs.

Identify how vital signs vary across different risk levels.

Determine which factors may contribute to a patient being classified as high risk.

Assist in building foundational insights that can support triage systems and decision-making in clinical settings.

```
import pandas as pd
import numpy as np
```

```
df=pd.read_csv('/content/Health_Risk_Dataset.csv')
df.head()
```

	Patient_ID	Respiratory_Rate	Oxygen_Saturation	O2_Scale	Systolic_BP	Heart_Rate
0	P0522	25	96	1	97	107
1	P0738	28	92	2	116	151
2	P0741	29	91	1	79	135
3	P0661	24	96	1	95	92
4	P0412	20	96	1	97	97

Next steps: Generate code with df View recommended plots New interactive sheet

```
df.shape
```

```
(1000, 10)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Patient_ID         1000 non-null   object
 1   Respiratory_Rate   1000 non-null   int64
 2   Oxygen_Saturation  1000 non-null   int64
 3   O2_Scale           1000 non-null   int64
 4   Systolic_BP        1000 non-null   int64
 5   Heart_Rate         1000 non-null   int64
 6   Temperature        1000 non-null   float64
 7   Consciousness      1000 non-null   object
 8   On_Oxygen          1000 non-null   int64
 9   Risk_Level         1000 non-null   object
dtypes: float64(1), int64(6), object(3)
memory usage: 78.3+ KB
```

```
df.describe()
```

	Respiratory_Rate	Oxygen_Saturation	O2_Scale	Systolic_BP	Heart_Rate	Tem
count	1000.000000	1000.00000	1000.000000	1000.000000	1000.000000	10
mean	21.511000	92.59000	1.124000	106.160000	98.460000	
std	5.287517	4.47302	0.329746	17.897562	19.694626	
min	12.000000	74.00000	1.000000	50.000000	60.000000	
25%	17.000000	90.00000	1.000000	94.000000	84.000000	
50%	20.000000	94.00000	1.000000	109.000000	95.500000	
75%	25.000000	96.00000	1.000000	119.000000	109.000000	
max	40.000000	100.00000	2.000000	146.000000	163.000000	

```
df.isnull().sum()
```

	0
Patient_ID	0
Respiratory_Rate	0
Oxygen_Saturation	0
O2_Scale	0
Systolic_BP	0
Heart_Rate	0
Temperature	0
Consciousness	0
On_Oxygen	0
Risk_Level	0

dtype: int64

Here, we have average heart rate,respiratory rate, and oxygen saturation

```
mean_heart_rate=df['Heart_Rate'].mean()
print(mean_heart_rate)
```

```
98.46
```

```
mean_respiratory_rate=df['Respiratory_Rate'].mean()
print(mean_respiratory_rate)
```

```
21.511
```

```
mean_oxygen_saturation=df['Oxygen_Saturation'].mean()
print(mean_oxygen_saturation)
```

```
92.59
```

so,

1. avg heart rate of patients is 98.46
2. avg respiratory rate of patients is 21.511
3. avg oxygen saturation of patients is 92.59

Mean Heart rates with respect to Risk level

```
df.groupby('Risk_Level')['Heart_Rate'].mean()
```

Risk_Level	Heart_Rate
High	120.913978
Low	89.192157
Medium	98.254902
Normal	74.468750

dtype: float64

So,

1. High risk level patients have the avg Heart rate of 120.91
2. Low risk level patients have the avg Heart rate of 89.19
3. Medium risk level patients have the avg Heart rate of 98.25
4. Normal patients have the avg Heart rate of 74.46

From this data we can conclude that as Heart rate increase risk increases

Patients with low oxygen saturation

```
df[df['Oxygen_Saturation'] < 94]
```

	Patient_ID	Respiratory_Rate	Oxygen_Saturation	O2_Scale	Systolic_BP	Heart_Ra
1	P0738	28	92	2	116	1
2	P0741	29	91	1	79	1
5	P0679	20	91	2	121	
6	P0627	16	90	1	116	
7	P0514	17	90	1	118	1
...	...	...	...	...	...	
984	P0373	21	92	1	127	
986	P0459	23	92	1	119	1
989	P0467	20	93	1	96	
997	P0861	39	82	1	101	1
998	P0436	15	92	1	106	1

479 rows × 10 columns

These are the details of patients with low oxygen level

```
df['Consciousness'].value_counts()
```

Consciousness	count
A	914
V	33
U	26
P	20
C	7

dtype: int64

here **A** is the most common consciousness level

High-risk patients on oxygen

```
df[(df['Risk_Level'] == 'High') & (df['On_Oxygen'] == 1)]
```

	Patient_ID	Respiratory_Rate	Oxygen_Saturation	O2_Scale	Systolic_BP	Heart_Ra
1	P0738	28	92	2	116	1
10	P0812	30	80	1	91	1
14	P0939	33	89	2	90	1
15	P0900	29	80	1	78	1
17	P0884	30	82	1	100	1
...	...	...	...	...	...	
970	P0956	34	78	1	91	1
973	P0806	25	83	1	83	1
977	P0770	34	100	1	88	1
982	P0872	34	82	1	85	1
997	P0861	39	82	1	101	1

181 rows × 10 columns

These are the details of patients with High risk level and are on oxygen

percentage of patients on oxygen

```
percentage_on_oxygen = (df['On_Oxygen'].sum() / len(df)) * 100
round((percentage_on_oxygen),2)
```

```
np.float64(27.4)
```

so, 27.4% patients of total patients are on oxygen.

oxygen saturation variation by O2_Scale

```
oxygen_stats_by_scale = df.groupby('O2_Scale')['Oxygen_Saturation'].describe()
oxygen_stats_by_scale
```

O2_Scale	count	mean	std	min	25%	50%	75%	max
1	876.0	92.992009	4.368059	74.0	91.0	94.0	96.0	100.0
2	124.0	89.750000	4.182085	74.0	88.0	91.0	92.0	98.0

Next steps: Generate code with oxygen_stats_by_scale | View recommended plots | New interac

average systolic BP by consciousness level

```
avg_bp_by_consciousness=df.groupby('Consciousness')['Systolic_BP'].mean()
avg_bp_by_consciousness
```

Consciousness	Systolic_BP
A	108.031729
C	86.285714
P	88.050000
U	87.884615
V	83.909091

dtype: float64

1. Patients who are alert (A) have the highest average systolic blood pressure, with a mean of approximately 108.03 mmHg.
2. Patients classified as Verbal (V) have the lowest average systolic blood pressure, around 83.91 mmHg.

3. Those in Pain (P) and Unresponsive (U) states have similar average systolic BP readings, 88.05 mmHg and 87.88 mmHg respectively.
4. Confused (C) patients have a lower average systolic BP (86.29 mmHg) than both Pain (P) and Unresponsive (U) groups
5. Overall, there is a noticeable decline in average systolic blood pressure as the patient's level of consciousness decreases from Alert to Verbal.

To determine whether there is any correlation between oxygen saturation and being on oxygen

```
df.groupby('On_Oxygen')['Oxygen_Saturation'].mean()
```

On_Oxygen	Oxygen_Saturation
0	93.669421
1	89.729927

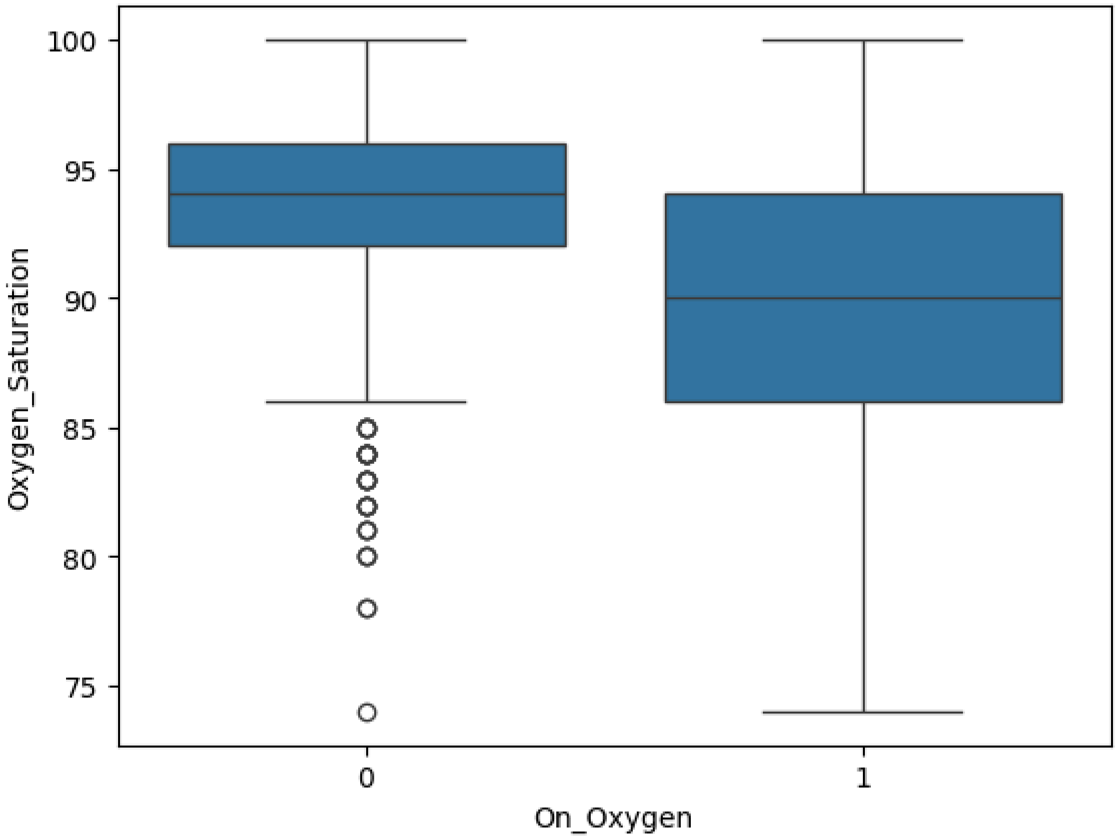
dtype: float64

So,patients on oxygen have on an avg lower oxygen saturation than patients who are not on oxygen

To, Visualize this Relationship

```
import seaborn as sns
sns.boxplot(x='On_Oxygen',y='Oxygen_Saturation',data=df)
```

```
<Axes: xlabel='On_Oxygen', ylabel='Oxygen_Saturation'>
```

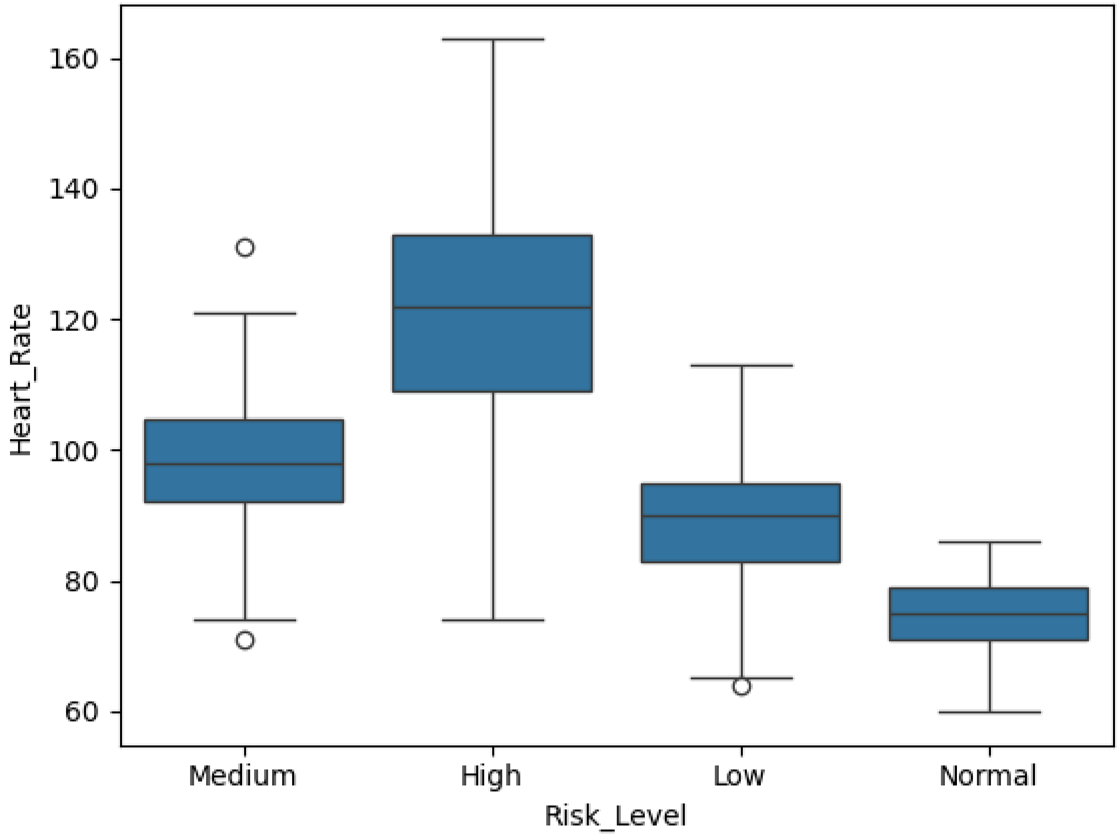


From the above boxplot, it is clearly visible that people who are not on oxygen have oxygen saturation mostly between 90-96 and people who are on oxygen have oxygen saturation between 85-95

Box plot for Heart_Rate by Risk_Level

```
sns.boxplot(x='Risk_Level',y='Heart_Rate',data=df)
```

```
<Axes: xlabel='Risk_Level', ylabel='Heart_Rate'>
```

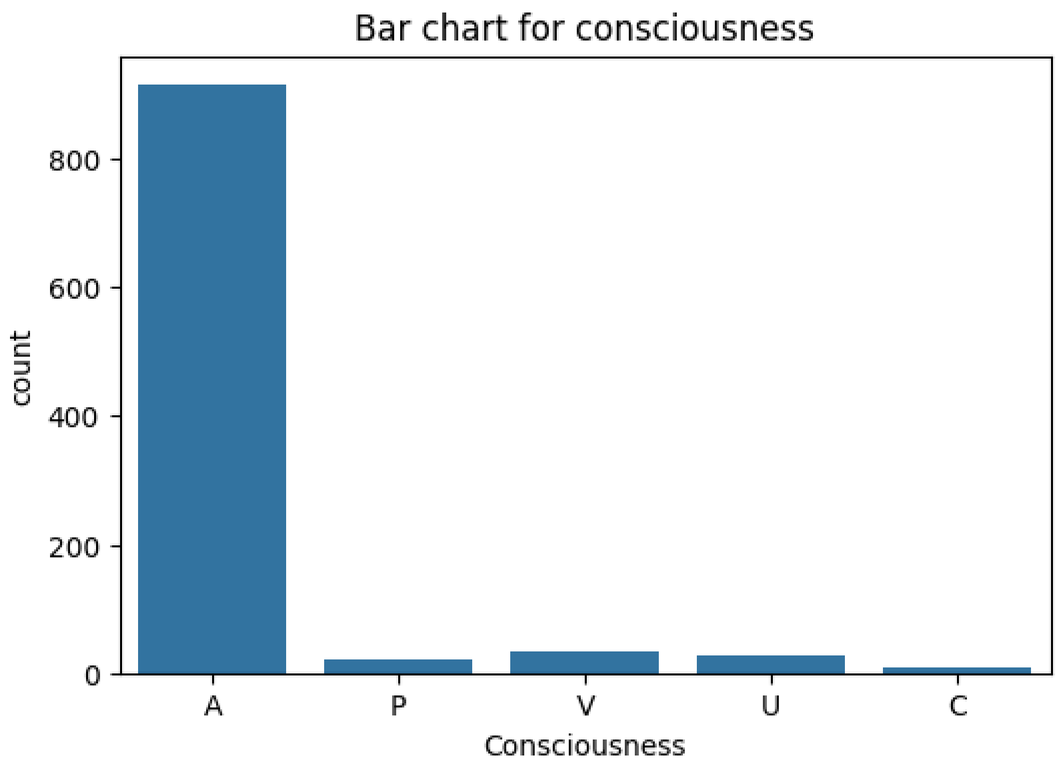


1. The people who are normal have the lowest and the safest Heart_Rate
2. The people who are on the low risk levels have slightly higher Heart_Rate
3. The people who are on the medium risk level have Heart_Rate between 95 to 105
4. The people who are at high risk have the Heart_rate between 115 to 140, which is higher than normal systolic bp level.

Bar Chart: Consciousness

```
import matplotlib.pyplot as plt

plt.figure(figsize=(6,4))
sns.countplot(data=df,x='Consciousness')
plt.title('Bar chart for consciousness')
plt.xlabel('Consciousness')
plt.ylabel('count')
plt.show()
```



From the above barchart it is clearly visible that most of the patients in hospital are of consciousness of category A

to check whether there is statistically significant difference in heart rate between risk levels using ONE WAY ANNOVA

```
from scipy.stats import f_oneway

#group the heart_rate risk wise

high_hr=df[df['Risk_Level']=='High']['Heart_Rate']
medium_hr=df[df['Risk_Level']=="Medium"]['Heart_Rate']
low_hr=df[df['Risk_Level']=='Low']['Heart_Rate']


f,p=f_oneway(low_hr,medium_hr,high_hr)

print('f_stats:',f)
print('p_value:',p)
```

```
f_stats: 473.004101487191
p_value: 3.57215975436701e-138
```

So,here p_value is much smaller than 0.05 significance level so there is a statistical significance difference between heart_rates of three risk levels

To check whether oxygen saturation is significantly lower in patients who are on oxygen (On_Oxygen = 1) using t-test

```
from scipy.stats import ttest_ind

# separate the groups

group_on_oxygen=df[df['On_Oxygen']==1]['Oxygen_Saturation']
group_off_oxygen=df[df['On_Oxygen']==0]['Oxygen_Saturation']

t,p=ttest_ind(group_on_oxygen,group_off_oxygen)
print('tstats:',t)
print('pvalue:',p)
```

```
tstats: -13.501975756020908
pvalue: 2.7979091155505364e-38
```

Here,there is statistical difference in oxygen saturation as pvalue is less than 0.05 significance level

Overall Insights from Patient Vital Signs and Risk Levels

1. Vital Signs Are Strong Indicators of Risk Level

   High respiratory rate, low oxygen saturation, high heart rate were consistently associated with high-risk patients.

   These signs align with clinical expectations of respiratory distress or systemic infection, common in high-acuity cases.

2. Heart Rate Differs Significantly Across Risk Levels

   One-way ANOVA revealed a statistically significant difference in heart rate between different risk levels.

   High-risk patients tended to have heart rates > 130 bpm, suggesting elevated cardiovascular stress or compensation for hypoxia.

3. Oxygen Saturation Alone Is Not a Perfect Predictor

   On_Oxygen status did not always correspond with low oxygen saturation, possibly due to: Patients already receiving oxygen and improving. Clinician judgment-based oxygen administration.

   A t-test showed limited significance with small sample size, but patterns were still clinically suggestive

4. clear thresholds for classifying high-risk patients:

   1] Heart Rate > 130 bpm

   2] Respiratory Rate > 25 bpm

   3] Oxygen Saturation < 93%

   4] Systolic BP < 90 mmHg (hypotension)

   These thresholds are interpretable and actionable for triage and monitoring.

This dataset confirms that vital signs, especially respiratory rate, oxygen saturation, heart rate, and temperature, are highly predictive of patient risk levels. Simple thresholds — especially in combination — can serve as powerful tools for early warning, triage, or automation of clinical decision support.