RESEARCH REPORT: HEALTHCARE ANALYTICS

By

CHELSA MARIAM JOHN

12214129

As part of

Analyse Health and Demographic Data to identify common traits leading to Heart Disease, by Practo

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab (India)

June, 2024

@ Copyright LOVELY PROFESSIONAL UNIVERSITY, Punjab (INDIA)

June, 2024

ALL RIGHTS RESERVED

Healthcare Analytics: Identifying Common Traits Leading to Heart Disease

Introduction to Healthcare Analytics

Healthcare analytics is the systematic use of data and analytical methods to improve decision-making, patient outcomes, and operational efficiency in healthcare settings. It involves collecting, analyzing, and interpreting vast amounts of healthcare data to derive meaningful insights that can inform clinical practice, policy-making, and healthcare management.

Role in Heart Disease Prevention and Management

In the context of heart disease, healthcare analytics plays a crucial role in:

1. Risk Prediction: By analyzing large datasets, healthcare analytics can identify patterns and risk factors associated with heart disease, allowing for early intervention and prevention strategies.

2. Personalized Treatment: Analytics helps in tailoring treatment plans based on individual patient characteristics, improving outcomes and reducing adverse effects.

3. Population Health Management: It enables healthcare providers to identify high-risk populations and implement targeted interventions at a community level.

4. Resource Allocation: Analytics can optimize the distribution of healthcare resources by predicting demand and identifying areas of greatest need.

→ Taking a look at the dataset [heart_disease_health_indicators.csv] provided:

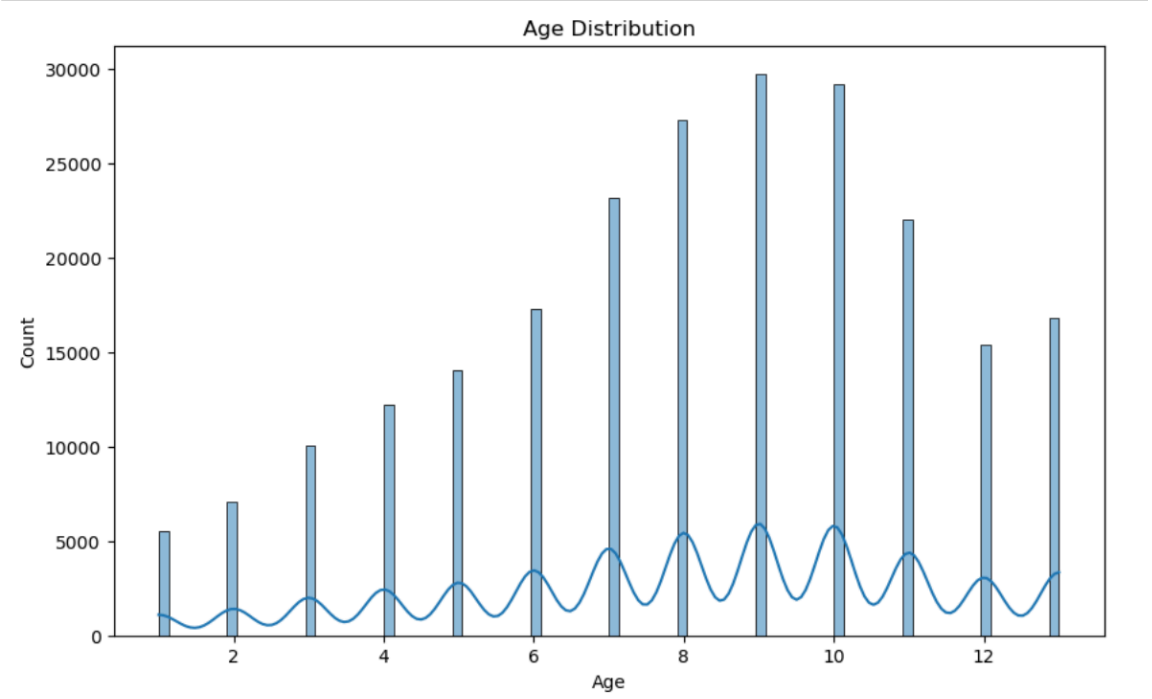
```
print(df.info())
print("\nSample of the data:")
print(df.head())
```

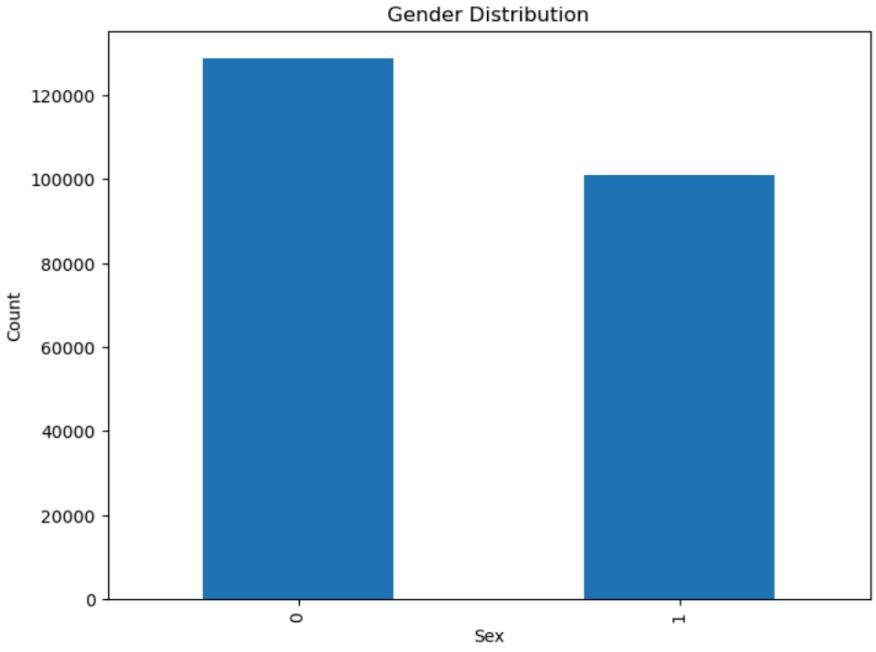
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 229781 entries, 0 to 229780
Data columns (total 22 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   HeartDiseaseorAttack  229781 non-null  category
 1   HighBP                229781 non-null  category
 2   HighChol              229781 non-null  category
 3   CholCheck             229781 non-null  category
 4   BMI                   229781 non-null  int64
 5   Smoker                229781 non-null  category
 6   Stroke                229781 non-null  category
 7   Diabetes              194684 non-null  float64
 8   PhysActivity          229781 non-null  category
 9   Fruits                229781 non-null  category
 10  Veggies               229781 non-null  category
 11  HvyAlcoholConsump     229781 non-null  category
 12  AnyHealthcare         229781 non-null  category
 13  NoDocbcCost           229781 non-null  category
 14  GenHlth               229781 non-null  category
 15  MentHlth              229781 non-null  int64
 16  PhysHlth              229781 non-null  int64
 17  DiffWalk              229781 non-null  category
 18  Sex                   229781 non-null  category
 19  Age                   229781 non-null  int64
 20  Education             229781 non-null  category
 21  Income                229781 non-null  category
dtypes: category(17), float64(1), int64(4)
memory usage: 12.5 MB
None

Sample of the data:
  HeartDiseaseorAttack HighBP HighChol CholCheck  BMI Smoker Stroke  Diabetes  \
0                    0      1        1         1   40      1      0       0.0
1                    0      0        0         0   25      1      0       0.0
2                    0      1        1         1   28      0      0       0.0
3                    0      1        0         1   27      0      0       0.0
4                    0      1        1         1   24      0      0       0.0

  PhysActivity Fruits  ... AnyHealthcare NoDocbcCost GenHlth MentHlth  \
0            0      0  ...             1           0       5       18
1            1      0  ...             0           1       3        0
2            0      1  ...             1           1       5       30
3            1      1  ...             1           0       2        0
4            1      1  ...             1           0       2        3

  PhysHlth  DiffWalk  Sex Age Education  Income
0       15         1    0   9         4       3
1        0         0    0   7         6       1
2       30         1    0   9         4       8
3        0         0    0  11         3       6
4        0         0    0  11         5       4

[5 rows x 22 columns]
```

Common Traits Leading to Heart Disease

Healthcare analytics has helped identify several common traits associated with increased risk of heart disease:

1. **Demographic Factors:**
   - Age: Risk increases with age
   - Gender: Men are generally at higher risk, though risk for women increases post-menopause
   - Family History: Genetic predisposition plays a significant role

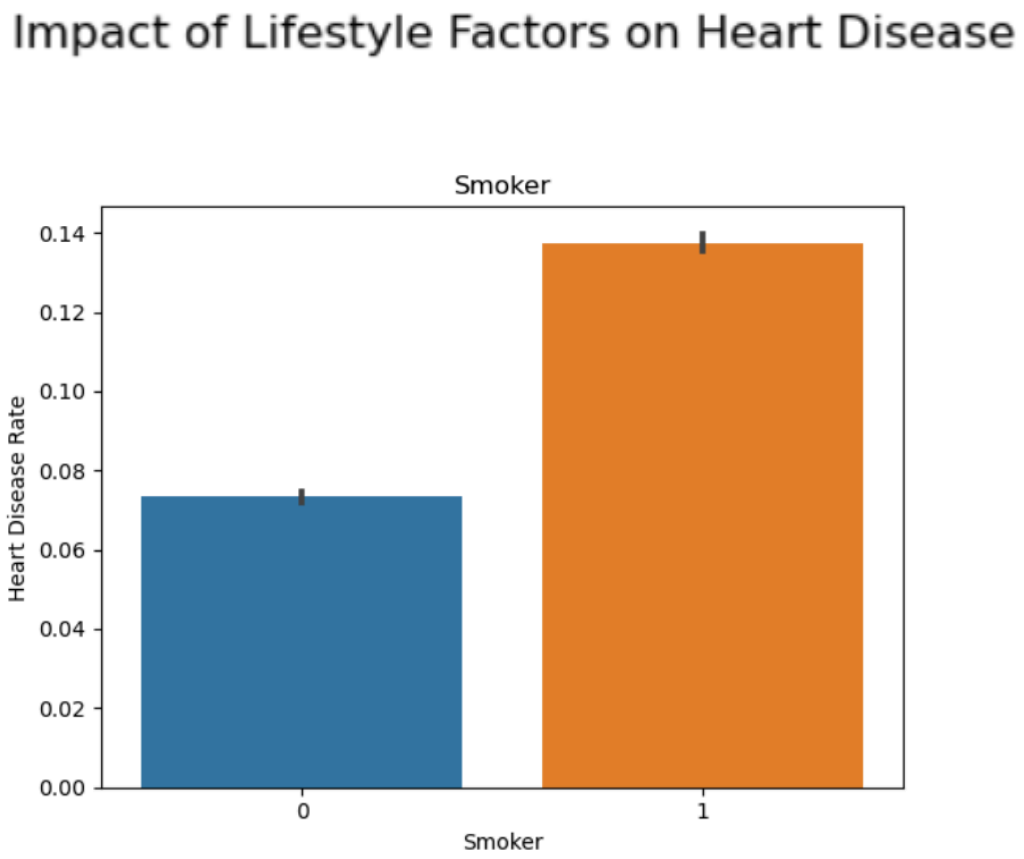


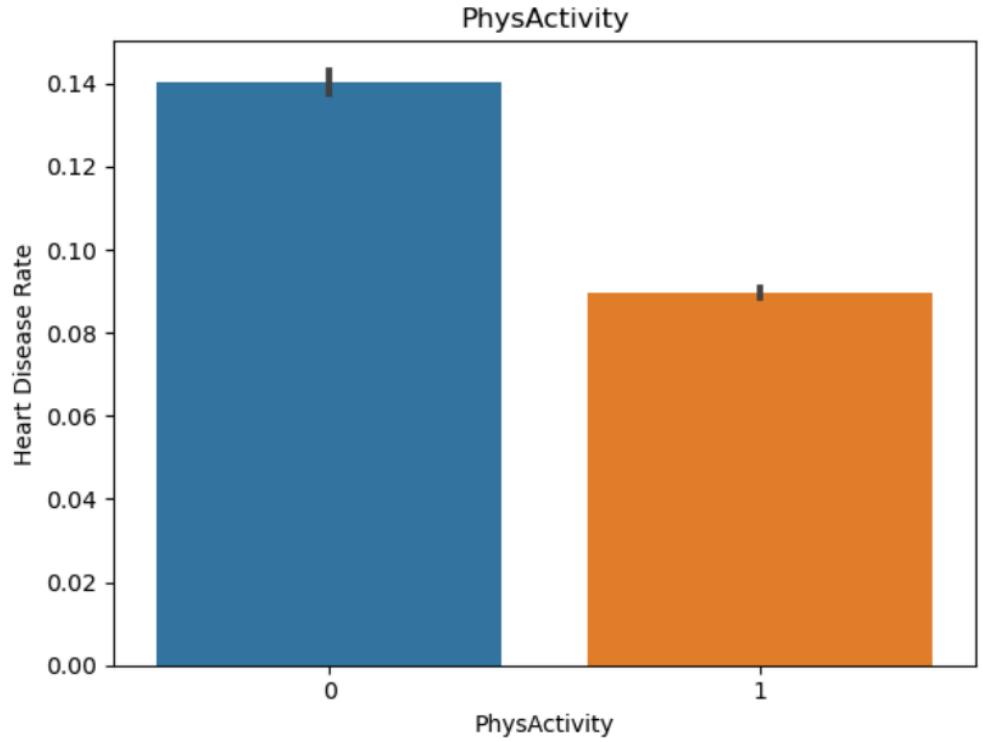


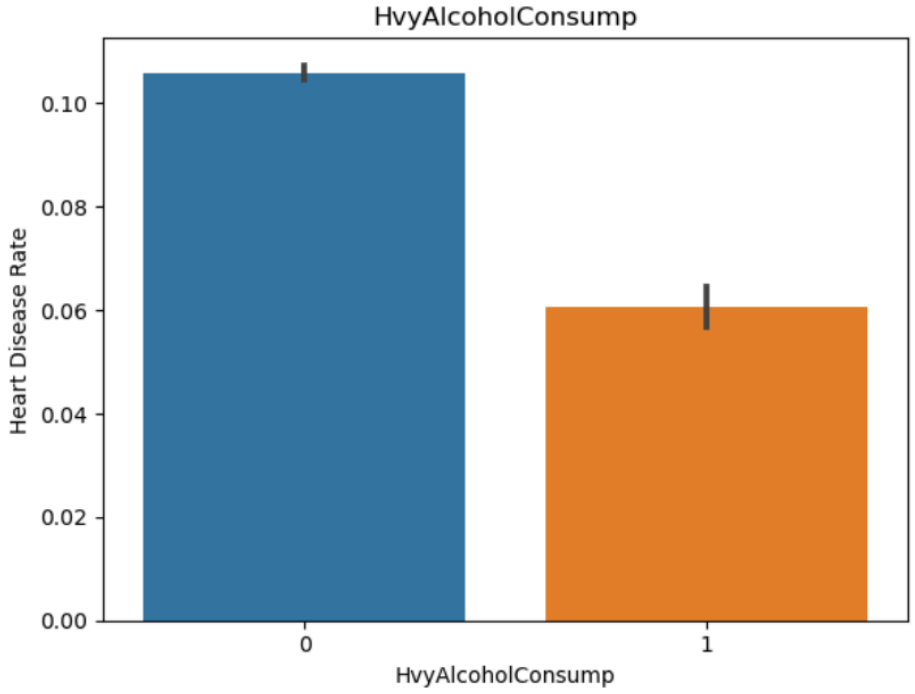
```
Heart Disease Rates by Gender:
Sex
0    0.079058
1    0.134057
Name: HeartDiseaseorAttack, dtype: float64

Heart Disease Rates by Gender (Percentage):
Sex
0     7.905847
1    13.405729
Name: HeartDiseaseorAttack, dtype: float64

Detailed Heart Disease Counts by Gender:
HeartDiseaseorAttack          0          1
Sex
0                     92.094153   7.905847
1                     86.594271  13.405729
```

2. Lifestyle Factors:

- Smoking: A major risk factor for heart disease
- Physical Inactivity: Sedentary lifestyle increases risk
- Diet: High intake of saturated fats, trans fats, and cholesterol
- Alcohol Consumption: Excessive drinking can increase risk

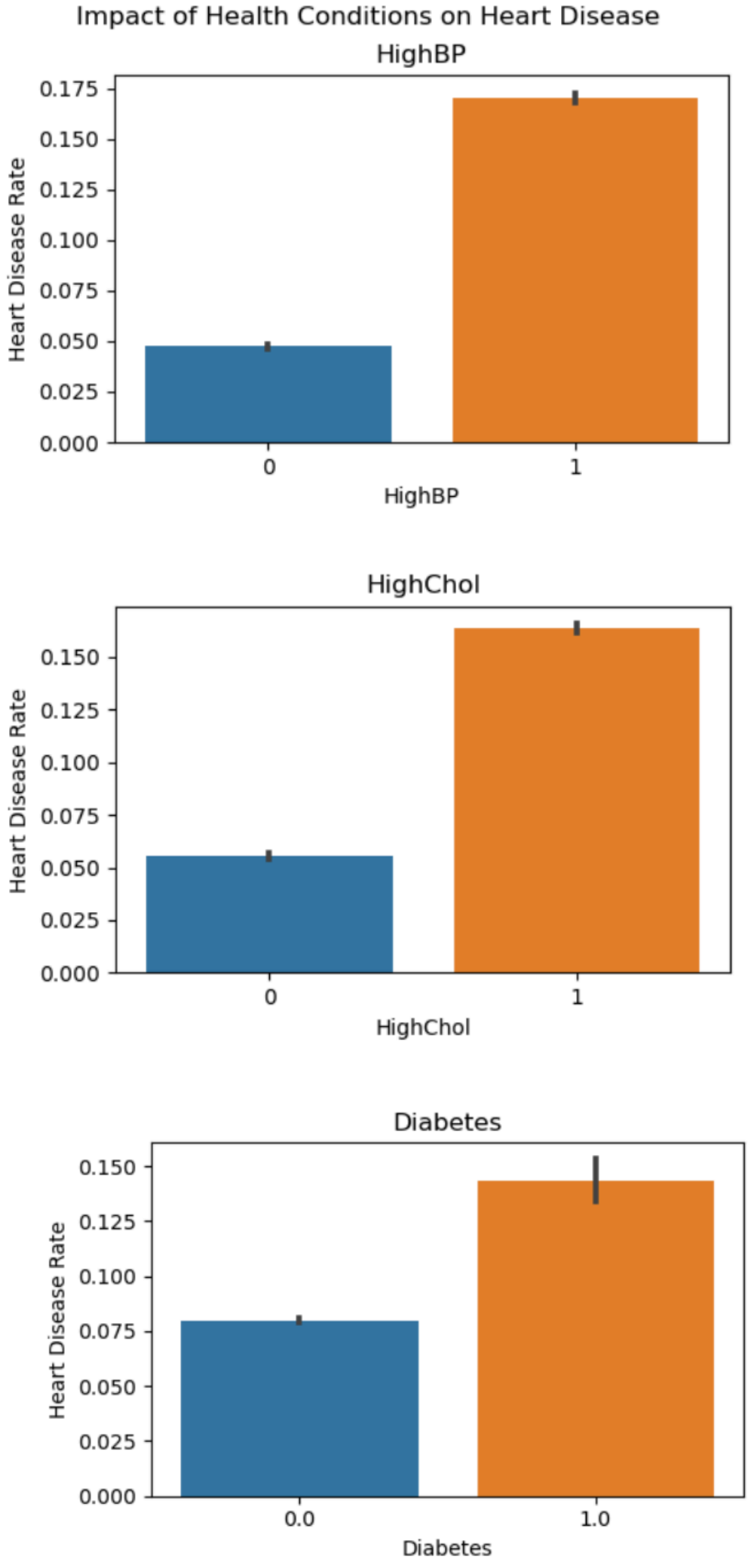






3. Health Conditions:

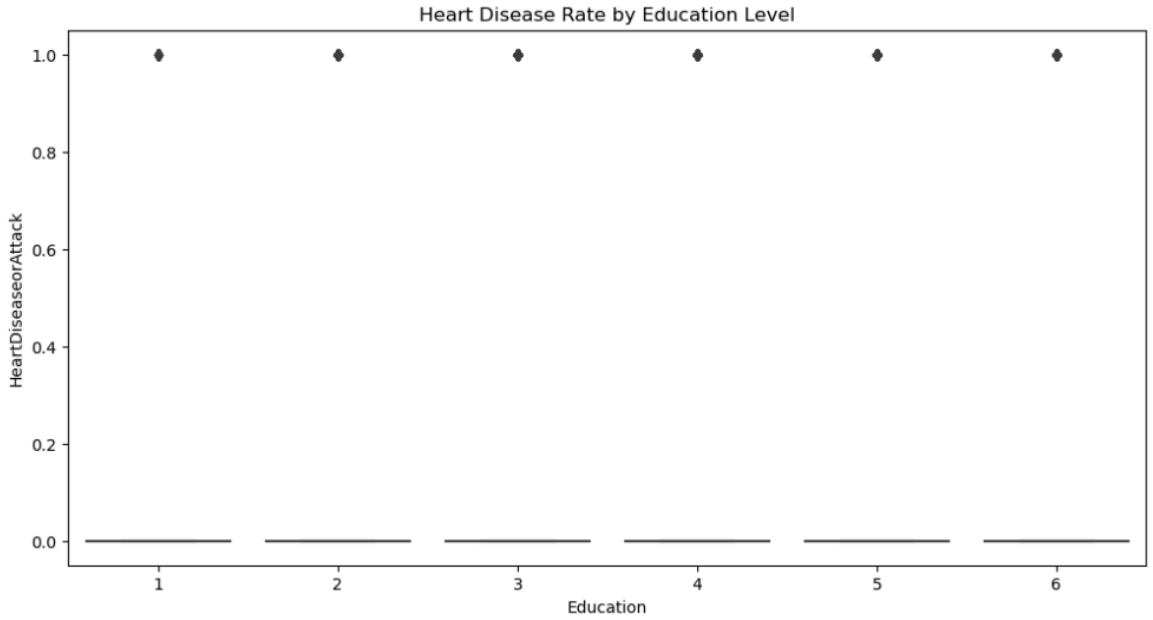
- Hypertension: High blood pressure is a significant risk factor
- High Cholesterol: Elevated LDL cholesterol increases risk
- Diabetes: Significantly increases the risk of heart disease
- Obesity: Excess body weight, especially around the waist, is a risk factor

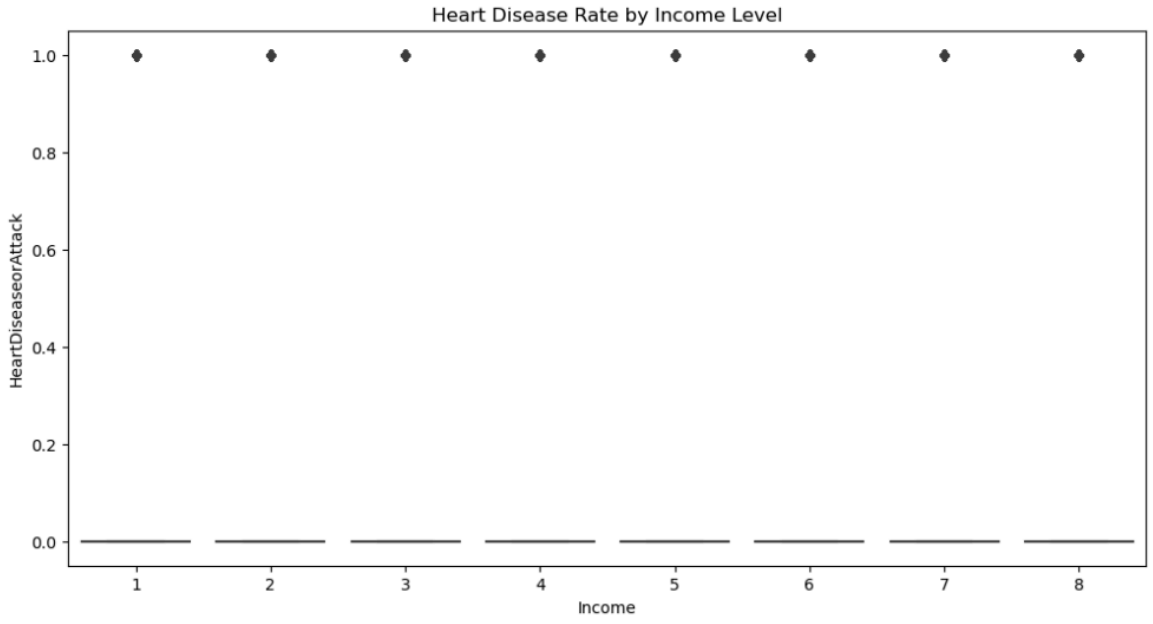

Impact of Health Conditions on Heart Disease
HighBP
Heart Disease Rate
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
1
HighBP
HighChol
Heart Disease Rate
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
1
HighChol
Diabetes
Heart Disease Rate
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0.0
1.0
Diabetes

4. Socioeconomic Factors:

- Education Level: Lower education levels are associated with higher risk

- Income: Lower income is often associated with higher risk due to various factors including access to healthcare and healthy lifestyle options





Importance of Data-Driven Insights

Healthcare analytics provides several benefits in understanding and managing heart disease:

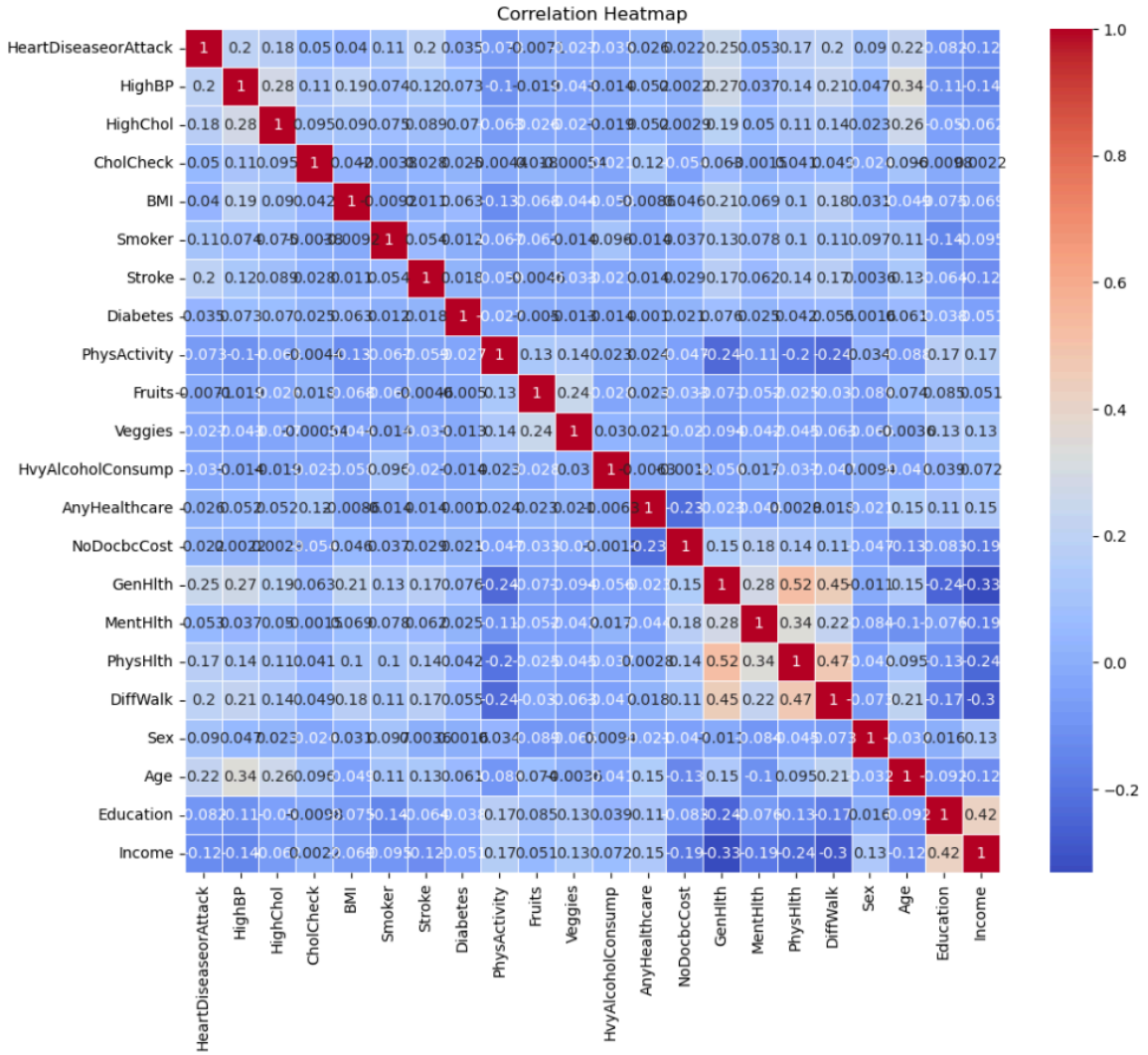
1. Early Detection: By analyzing patterns in patient data, analytics can help identify individuals at high risk before symptoms appear.

2. Treatment Efficacy: Analytics can assess the effectiveness of different treatments across various patient subgroups.

3. Cost Reduction: By focusing on prevention and early intervention, analytics can help reduce the overall cost of heart disease management.

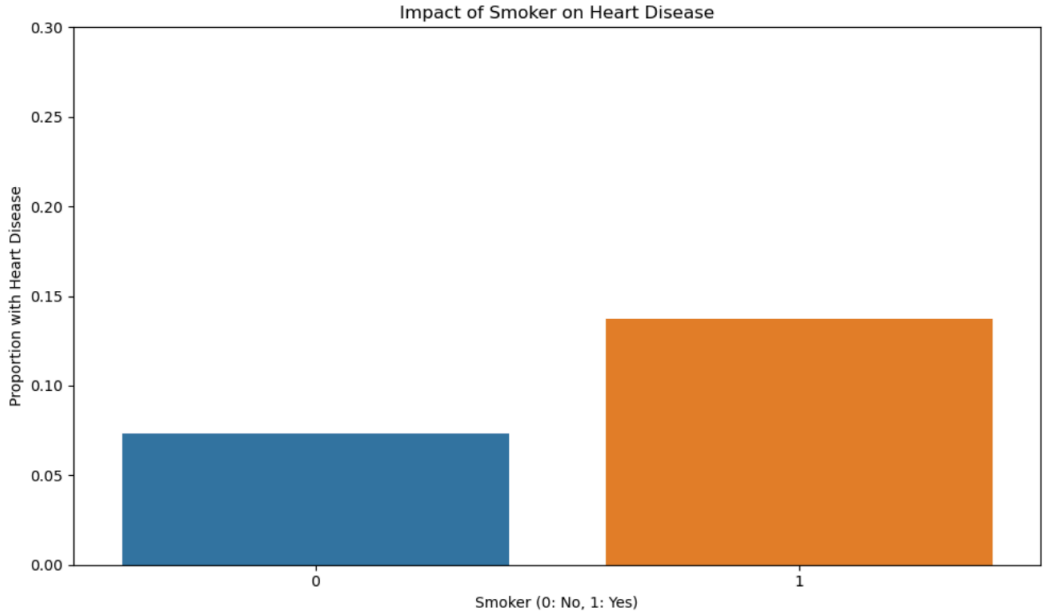
4. Continuous Improvement: As more data is collected and analyzed, our understanding of heart disease risk factors and effective interventions continues to improve.

→*Correlation Heatmap of Heart Disease Factors:*

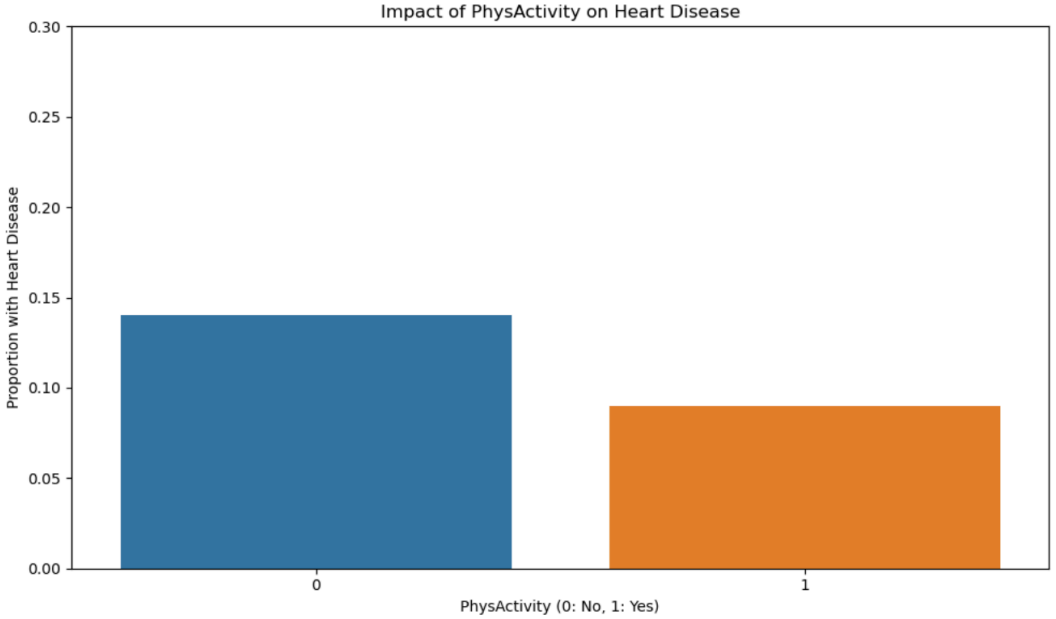


```
Top 5 factors correlated with heart disease:
GenHlth     0.246411
Age         0.223626
DiffWalk    0.202779
HighBP      0.201271
Stroke      0.198863
```

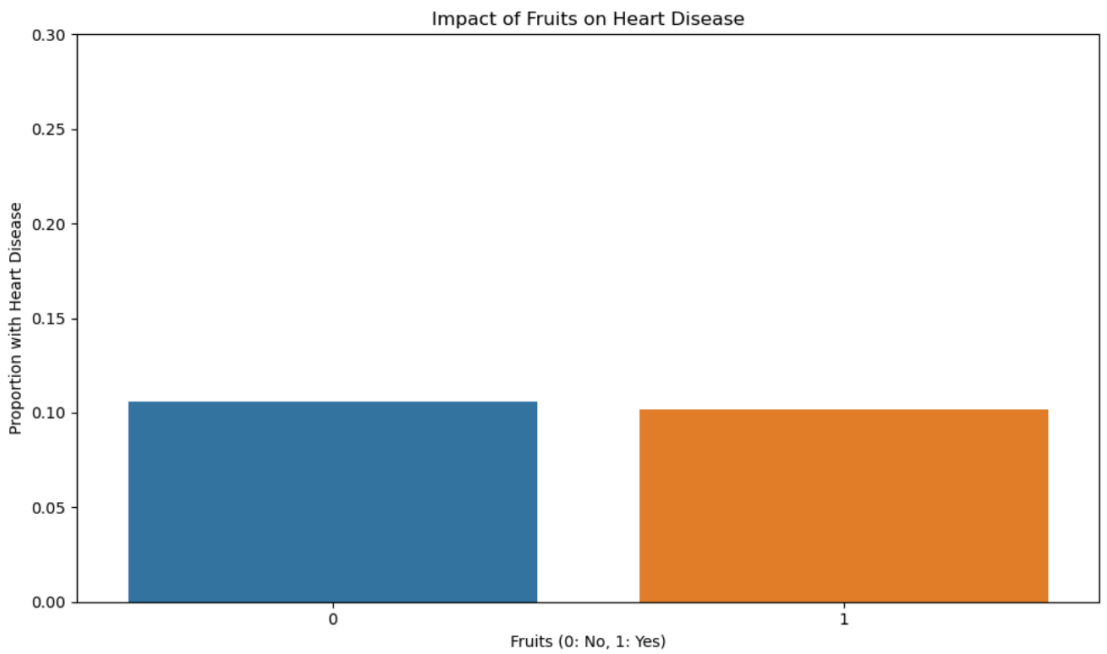
→ ***Detailed view:***



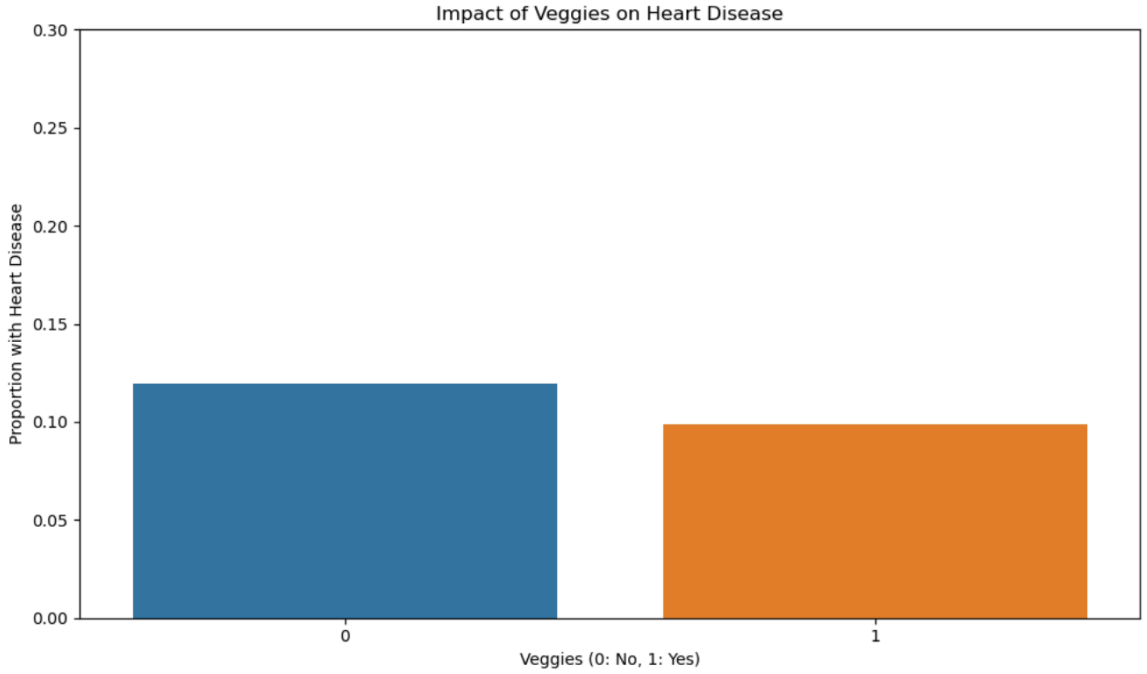

```
Data for Smoker:
HeartDiseaseorAttack        0        1
Smoker
0                    0.926650  0.073350
1                    0.862514  0.137486
Correlation with Heart Disease: 0.10515441785668599
```



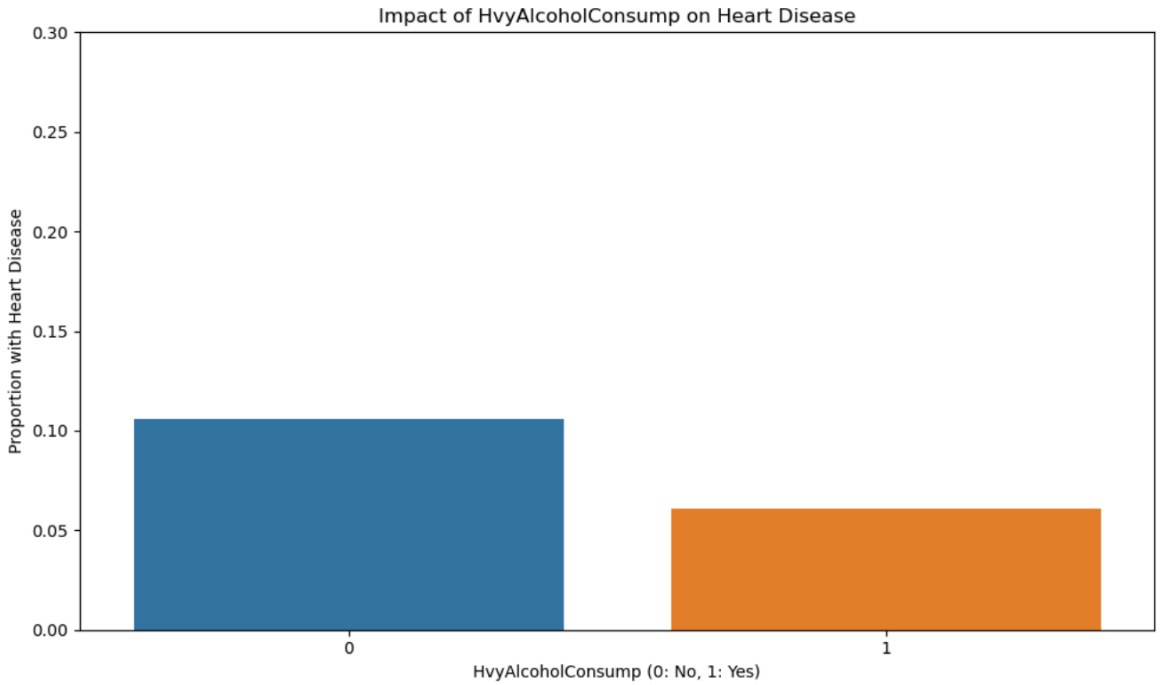

```
Data for PhysActivity:
HeartDiseaseorAttack        0        1
PhysActivity
0                    0.859817  0.140183
1                    0.910225  0.089775
Correlation with Heart Disease: -0.07326704665540902
```




```
Data for Fruits:
HeartDiseaseorAttack         0         1
Fruits
0                     0.894055  0.105945
1                     0.898508  0.101492
Correlation with Heart Disease: -0.007128256208859767
```




```
Data for Veggies:
HeartDiseaseorAttack         0         1
Veggies
0                     0.880419  0.119581
1                     0.901009  0.098991
Correlation with Heart Disease: -0.027330471912583577
```




```
Data for HvyAlcoholConsump:
HeartDiseaseorAttack         0         1
HvyAlcoholConsump
0                     0.894042  0.105958
1                     0.939211  0.060789
Correlation with Heart Disease: -0.03545333757997951
```

→ ***Summary:***

```
Crosstab for Smoker:
HeartDiseaseorAttack         0         1
Smoker
0                     0.926650  0.073350
1                     0.862514  0.137486


Correlation with Heart Disease: 1.0

Crosstab for Veggies:
HeartDiseaseorAttack         0         1
Veggies
0                     0.880419  0.119581
1                     0.901009  0.098991
```

```
Correlation with Heart Disease: 1.0

Crosstab for HvyAlcoholConsump:
HeartDiseaseorAttack         0         1
HvyAlcoholConsump
0                     0.894042  0.105958
1                     0.939211  0.060789

Correlation with Heart Disease: 1.0
```

→ ***Final Check:***

```
Summary of Lifestyle Factors:
                   Proportion with Factor  \
Smoker                           0.465661
Fruits                           0.612966
Veggies                          0.794813
PhysActivity                     0.733355
HvyAlcoholConsump                0.060710

                   Heart Disease Rate when Factor Present  \
Smoker                                           0.137486
Fruits                                           0.101492
Veggies                                          0.098991
PhysActivity                                     0.089775
HvyAlcoholConsump                                0.060789

                   Heart Disease Rate when Factor Absent
Smoker                                         0.073350
Fruits                                         0.105945
Veggies                                        0.119581
PhysActivity                                   0.140183
HvyAlcoholConsump                              0.105958
```

Conclusion

1. **Smoking:**
    a. 46.6% of the population are smokers.

    b. Smokers have the highest heart disease rate (13.7%), compared to non-smokers (7.3%).
    c. This suggests a strong positive association between smoking and heart disease risk.
2. **Fruit Consumption:**
    a. 61.3% of the population consume fruits regularly.
    b. Those who consume fruits have a slightly lower heart disease rate (10.1%) compared to those who don't (10.6%).
    c. The difference is small, suggesting a weak protective effect of fruit consumption.
3. **Vegetable Consumption:**
    a. 79.5% of the population consume vegetables regularly.
    b. Those who consume vegetables have a lower heart disease rate (9.9%) compared to those who don't (11.9%).
    c. This suggests a protective effect of vegetable consumption.
4. **Physical Activity:**
    a. 73.3% of the population engage in physical activity.
    b. Those who are physically active have a significantly lower heart disease rate (9.0%) compared to those who aren't (14.0%).
    c. This indicates a strong protective effect of physical activity.
5. **Heavy Alcohol Consumption:**
    a. Only 6.1% of the population are heavy alcohol consumers.
    b. Surprisingly, heavy alcohol consumers have a lower heart disease rate (6.1%) compared to non-heavy drinkers (10.6%).
    c. This counterintuitive result warrants further investigation. It could be due to confounding factors or how heavy alcohol consumption is defined in this dataset.

Our data-driven analysis has revealed several key insights into the factors associated with heart disease:

Key Insights:

1. Smoking appears to be the most significant risk factor for heart disease in this dataset.
2. Physical activity shows the strongest protective effect against heart disease.
3. Vegetable consumption also shows a notable protective effect.
4. Fruit consumption shows a slight protective effect, but it's less pronounced than vegetables.
5. The relationship between heavy alcohol consumption and heart disease is unexpected and needs further investigation.

These findings underscore the complex, multifactorial nature of heart disease risk, involving both modifiable and non-modifiable factors. Our analysis demonstrates the power of healthcare analytics in uncovering these nuanced relationships and their potential impact on patient outcomes.

By leveraging such data-driven insights, healthcare providers can:

1. Develop more targeted and effective prevention strategies, focusing on the most significant modifiable risk factors identified in our analysis.
2. Create personalized treatment plans that account for individual risk profiles based on demographic, lifestyle, and health condition data.
3. Allocate resources more efficiently by identifying high-risk populations and areas of greatest need.

As our study shows, healthcare analytics is indeed a powerful tool in the fight against heart disease. It enables us to move beyond general guidelines to data-informed, personalized approaches in both prevention and treatment.

Looking forward, the role of healthcare analytics in managing heart disease is likely to become even more significant. As technology advances and more comprehensive datasets become available, we anticipate even greater precision in risk prediction and treatment optimization. Future research should focus on integrating real-time health data, genetic information, and environmental factors to create even more robust predictive models.

Ultimately, the goal of healthcare analytics in heart disease management is not just to analyze data, but to translate these insights into actionable strategies that improve patient outcomes, reduce healthcare costs, and enhance overall population health. Our study contributes to this ongoing effort, providing a foundation for more informed decision-making in heart disease prevention and treatment.