page2image568

Data Visualization for Business Narration

Course-End Project Solution



**Exploratory Data Analysis**

1. **Visually explore the distribution of heart disease amongst individuals**

**Steps:**

1. Load the dataset ***heart.csv*** and the required packages using the following code:



1. Observe the output received:

Table

Description automatically generated

1. Visually explore the distribution of heart disease amongst individuals.

**Steps:**

1. For exploration, get frequency distribution data using table().
2. Compute percentage for labels.
3. Plot a pie chart using ggplot.



1. Observe the output below:

Chart, pie chart

Description automatically generated

1. **Explore the distribution of cases with Yes/No to heart disease according to sex**

**Hint:** Create a frequency table for Sex vs. Heart Disease.



Observe the output below:

Chart, bar chart

Description automatically generated

**Observation:** Males have a higher chance of heart disease as compared to females.

1. **Distribution of cases with Yes/No to heart disease according to whether the patient is a smoker or not**



Observe the output received:

Chart, bar chart

Description automatically generated

**Observation:** Non-smokers seem to have higher chances of heart disease.

1. **Distribution of Body Mass Index according to Yes/No to heart disease**

**Steps:**

1. Explore BMI using the following code:



Observe the output below:

Chart, histogram

Description automatically generated

1. Explore BMI vs. Heart Disease using the following code:



Observe the output received:

Chart, histogram

Description automatically generated

1. **Explore BMI vs. heart disease using a boxplot to identify outliers**

**Hint:** Boxplot provides information about skewness.

Execute the following code:



Observe the output received:

Chart, box and whisker chart

Description automatically generated

1. **Explore the general health variable across heart disease**

**Hint:** Explore using a bar chart.

**Steps:**

1. Explore the general health variable.



Observe the output below:

Chart, bar chart

Description automatically generated

1. Explore General Health vs. Heart Disease using the following code:



Observe the output received:

Chart, bar chart

Description automatically generated

1. **Analyze if sleep time affects the occurrence of heart disease**

**Hint:** Filter the data for only the individuals with the disease and study the sleeping time. Use the following code:



Text

Description automatically generated

Observe the output received:

Chart, histogram

Description automatically generated

**Observation:** Individuals sleeping between 6 to 8 hrs have higher chances of heart disease. This may also be because of the average sleep time in the overall data.