**Jupyter Notebook Homework 1.1:**

**Creating Data Visualizations**

**Due: 04th March 2025**

|  |  |  |  |
| --- | --- | --- | --- |
| **Individual Contribution** | | | |
| **CWID** | **Name** | **Contribution (description)** | **Percent**  **Contribution** |
| **A20563423** | **Ziwen Chen** | **Integrate and improve code, write report** | **33.3%** |
| **A20563465** | **Tianyi Tang** | **edit code, make visualizations** | **33.3%** |
| **A20563424** | **Licheng Rao** | **edit code, make visualizations** | **33.3%** |
|  |  |  |  |

Introduction of the dataset we use

This dataset examines the relationship between safety equipment and crash survival rates, including variables such as gender, helmet use, seatbelt use, airbag deployment, survival rate, and driving speed. By analyzing these variables, the impact of safety equipment such as helmets and seatbelts on survival rates can be assessed, providing data support for the formulation of traffic safety policies and public safety education.

Data processing

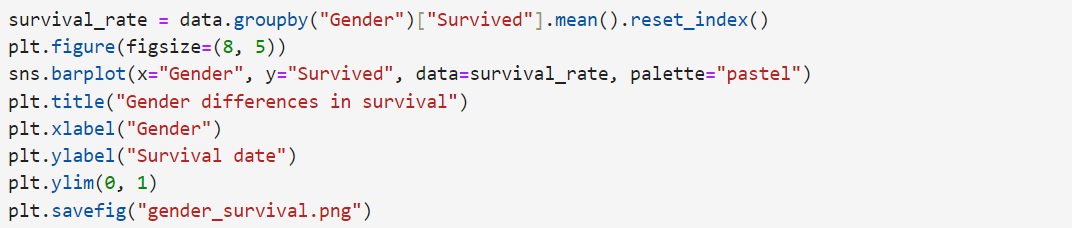
1. **Bar chart of gender and survival rate**

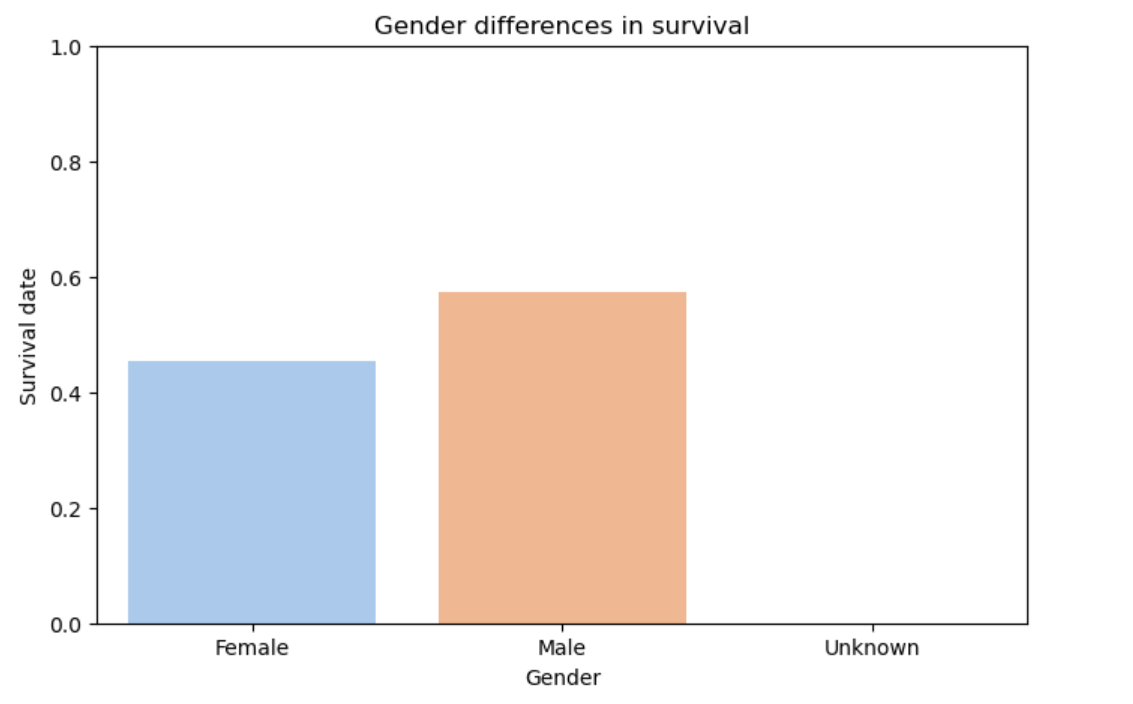
·Group by gender.

·Calculate the average survival rate for the two groups.

·Convert the results into a DataFrame.

·Use Seaborn to plot a bar chart.



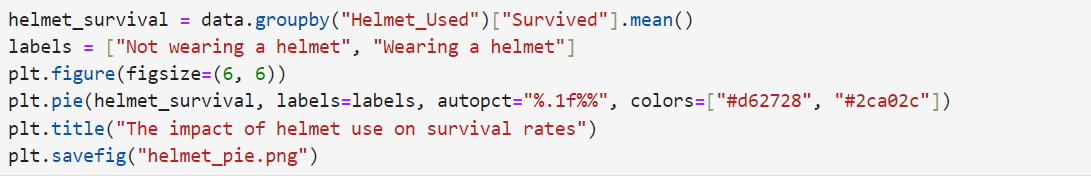


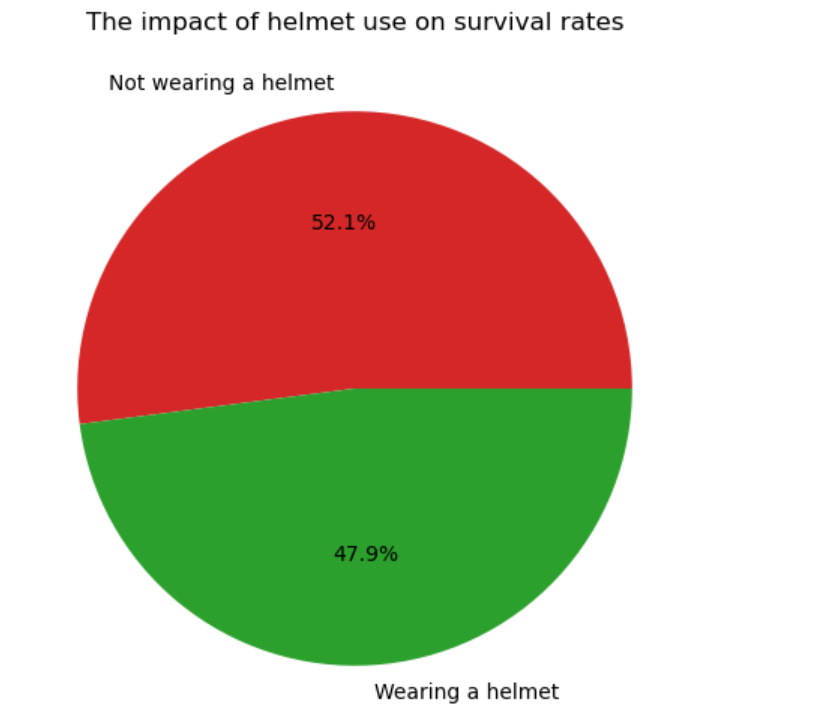
1. **Pie chart of survival rate based on helmet usage**

·Group by Helmet\_Used.

·Calculate the average survival rate for each group.

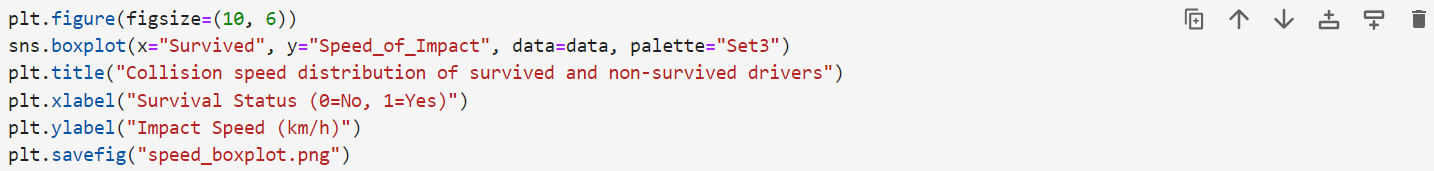
·Use the Matplotlib library to create a pie chart.

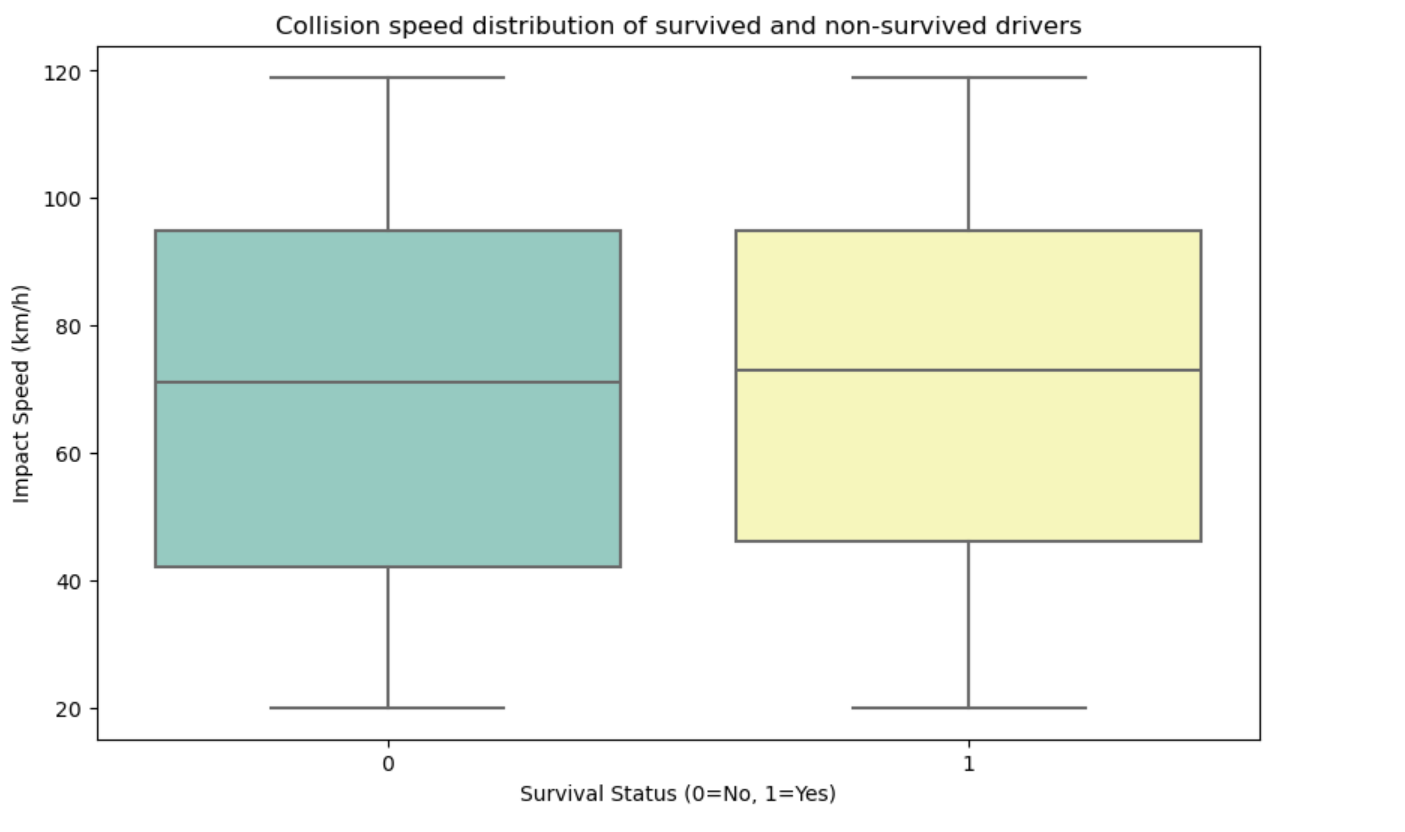




1. **Box plot of survival status and collision speed**  
   ·Use sns.boxplot to automatically group and plot box plots for Survived = 0 and Survived = 1.

·Label the x-axis as "Survival Status" and the y-axis as "Impact Speed."





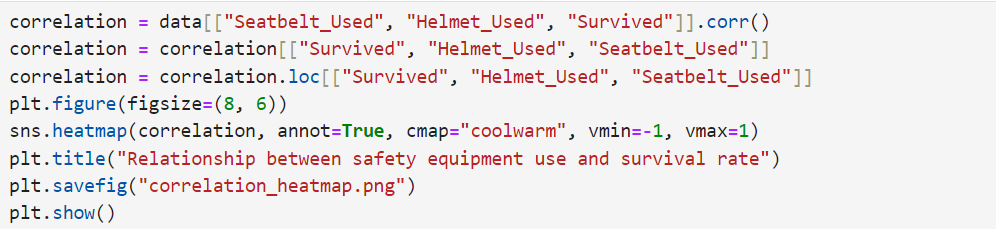
1. **Heatmap of safety equipment usage and survival rate**

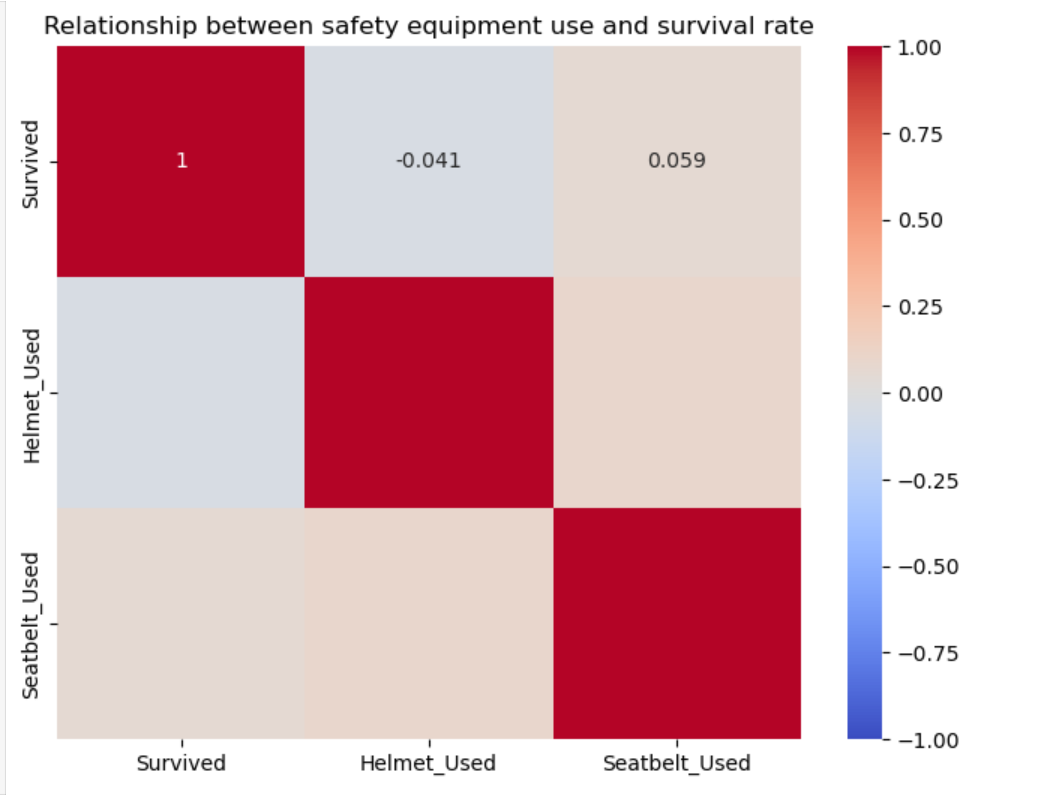
·Select the relevant columns (survival status, helmet usage, seatbelt usage).

·Calculate the correlation matrix for the selected columns.

·The correlation coefficient ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 representing no correlation.

·Use sns.heatmap to visualize the correlation between the variables, with high correlation shown in red and low correlation in blue.





The libraries we have used

·**pandas**：Read the CSV file, handle missing values, and perform data analysis.

·**matplotlib**：Create various charts.

·**seaborn：**Conduct advanced data visualization, primarily for creating bar charts, box plots, and heatmaps, to enhance the aesthetics of the charts.

Dataset analysis

**1. Gender and Survival Rate**

From the bar chart, it can be observed that the survival rate for males is higher than that for females. However, overall, both male and female survival rates are between 0.4 and 0.6, indicating that the difference is not significant. This may suggest that gender is not the primary factor determining survival rates, although it still has some influence.

**2. Helmet Usage and Survival Rate**

According to the pie chart, the survival rate for individuals wearing helmets is 47.9%, while the survival rate for those not wearing helmets is 52.1%. This result implies that wearing a helmet does not significantly increase the survival rate and may even be slightly lower.

**3. Collision Speed and Survival Rate**

The box plot shows that the collision speed range for non-survivors is between 40-100 km/h. Survivors also have collision speeds within the same range, but their speeds are more concentrated. This may indicate that higher collision speeds decrease the likelihood of survival, while survivors tend to have speeds concentrated within a certain range, possibly because excessively high speeds lead to fatal accidents, whereas lower speeds increase the chances of survival.

**4. Correlation Between Helmet/Seatbelt Usage and Survival Rate**

From the correlation matrix, the correlation between survival rate and helmet usage is -0.041 (a negative correlation, but with almost no impact). The correlation between survival rate and seatbelt usage is 0.059 (a positive correlation, but with minimal effect). This suggests that the relationship between helmet usage and survival rate is weak, and may even be slightly negative. The survival rate is slightly higher for seatbelt users, but the effect is not significant.

**Overall Conclusion**

1.Gender has a certain influence on survival rates, with males having a slightly higher survival rate, but the overall difference is not substantial.

2.The relationship between helmet usage and survival rate is weak, and may even be slightly negative, warranting further analysis.

3.Collision speed affects survival rates, but the speed ranges for non-survivors and survivors are similar, indicating that other factors may also need to be considered.

4.Seatbelt usage shows a slight positive correlation with survival rate, but the impact is not significant.