Titanic Dataset Analysis - Part2

Dataset Download

Download the Titanic dataset from:

- Kaggle: https://www.kaggle.com/c/titanic/data
- Direct CSV link:
  https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv

Assignment Overview

In this assignment, you will analyze the Titanic passenger dataset to uncover patterns and insights about survival factors during the Titanic disaster. You'll practice statistical analysis, and data visualisation using pandas, matplotlib, and seaborn.

Prerequisites

- Python 3.x
- pandas
- numpy
- matplotlib
- seaborn

Assignment Tasks

1. Basic Analysis

Calculate and present:
 a) Survival rates by:

- Passenger class
- Gender
- Age category
- Deck
- Port of embarkation

b) Family size impact:

- Analyze survival rates based on family size

- Determine if traveling alone affected survival chances

2. Advanced Analysis

Perform the following analyses:
 a) Create pivot tables showing:

- Survival rates by class and gender
- Survival rates by age group and class
- Survival rates by deck and class

b) Statistical tests:

- Correlation between fare and survival
- Chi-square test for independence between categorical variables

3. Visualization

Create at least three meaningful visualizations:

- Survival distribution by passenger class and gender
- Age distribution of survivors vs non-survivors
- Fare distribution analysis
- Family size impact on survival

Deliverables

1. Python script (.py) or Jupyter notebook (.ipynb)
2. Brief report (PDF) containing:
    - Summary of findings
    - Key insights
    - Methodology explanation
    - Visualizations
    - Conclusions

Tips

- Start with basic data exploration
- Document your assumptions
- Use functions for reusable code

- Include error handling
- Comment your code adequately
- Test your code with different scenarios

Note

- All code should be properly documented
- Include error handling where appropriate
- Follow Python PEP 8 style guidelines
- Use efficient pandas operations
- Avoid duplicate code

Resources

- Pandas documentation: https://pandas.pydata.org/docs/
- Seaborn documentation: https://seaborn.pydata.org/
- Matplotlib documentation: https://matplotlib.org/
- Kaggle Titanic tutorials: https://www.kaggle.com/c/titanic/tutorials