



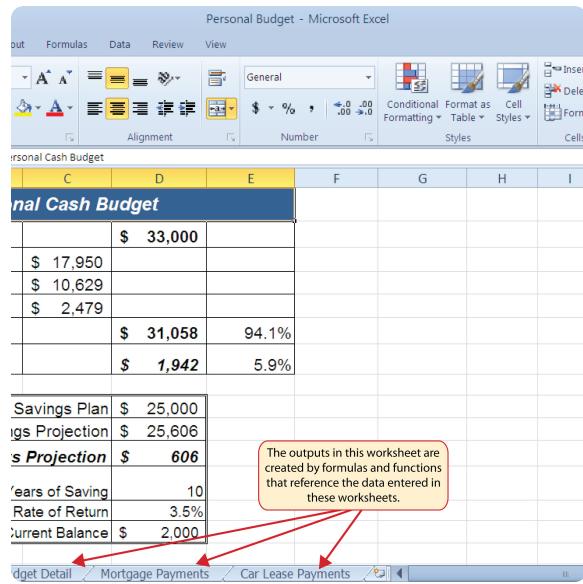
Titanic Survival Prediction

Explore the steps followed to create a model and predict survival for Titanic passengers on Kaggle.

Praveen Karthik Arumugam

www.linkedin.com/in/pk7779

Data Analysis



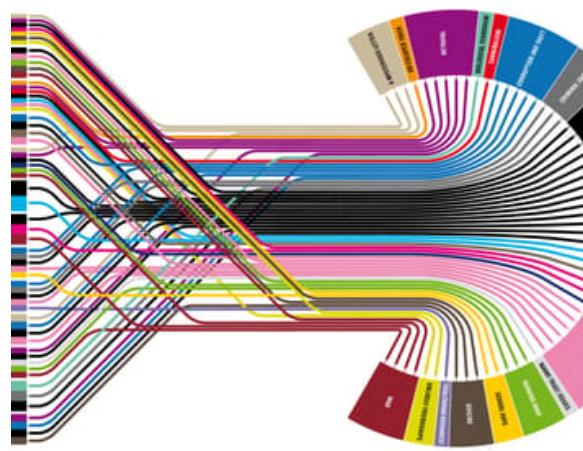
Data Analysis in Excel

Analyze the data to gain insights into a large dataset of Titanic passengers.



Data Cleaning in Python

Use Pandas, isna, Fillna, and Standard deviation to clean the data, preparing it for analysis and exploration.



Exploratory Data Analysis

Using visualization tools, represent data and explore relationships between different variables.



Feature Engineering

Encoding and selecting features to improve model accuracy.



Model Selection

1

Logistic Regression

Analyze relationships between features and output, modelling the probability of survival.

3

Random Forest

Combining multiple decision trees to improve performance and accuracy.

2

Decision Tree

Partitioning data into smaller groups based on different criteria.

4

Adaboost & Gradient Boost

Increase the accuracy of models by focusing on misclassified data points.

Feature Selection

Label Encoding

Convert categorical data into numerical values for easier analysis.

- Sex

SelectKBest

Select the most important features to improve model performance.

1. Age
2. SibSp
3. Pclass
4. Fare

Chi Square

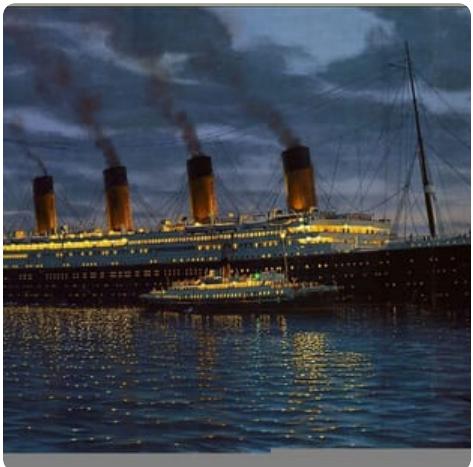
Identify the independencies and relationships between the variables.

- Age
- Fare
- Sex

Prediction

"A good model can give an unprecedented understanding of the relationships between different passenger attributes and help to identify the passengers who were more likely to survive the disaster."

- Kaggle Competition



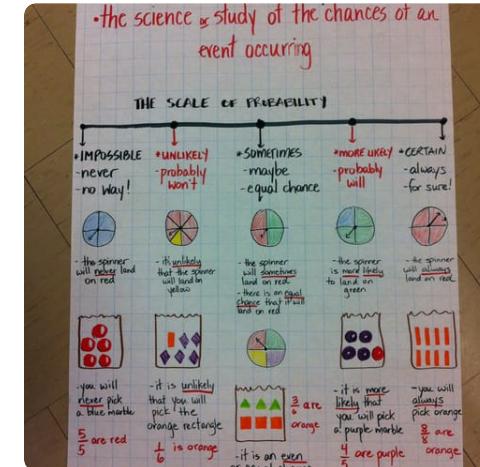
Passenger Data

Inputs: Age, Sex, Fare, and more.



Machine Learning Model

Model trains on the passenger data to predict the survival of the passenger based on the input attributes.



Output: Survival Probability

Percentage that a passenger survived or not.

Conclusion

Achievement

Success in predicting the survivors of the Titanic disaster.

New Learnings

Exploration of different data analysis and machine learning techniques.

Future Scope

Use machine learning models to analyze different data sets to gain insights and to create a better future.