Detailed Project Report (DPR)

Analysing world's best wine reviews

Document Version Control

Date Issue	Version	Description	Author
27/09/2024	1	Initial DPR - V 1.0	SAURAV
07/10/2024	2	Initial DPR - V 2.0	SAURAV

Project Overview

This project aims to analyze a dataset of wine reviews to extract valuable in sights and answer key questions related to wine quality, pricing, and other attributes. By leveraging data analytics techniques, we will explore patterns and relationships within the data to inform wine enthusiasts, producers, an d retailers.

Dataset Description

The dataset consists of the following columns:

1. **wine**: Name of the wine.
2. **winery**: Name of the winery.
3. **category**: Category of the wine (Red or White).
4. **designation**: Designation of the wine.
5. **varietal**: Type of grape used.
6. **appellation**: Geographic origin of the wine.
7. **alcohol**: Alcohol content percentage.
8. **price**: Price of the wine in USD.

9. **rating**: Rating of the wine.
10. **reviewer**: Name of the reviewer.
11. **review**: Text review of the wine.

Analysis Questions:

1. Number of Wines by Category:
    - Determine the distribution of red and white wines in the dataset.
2. Price by Category:
    - Analyze the average price of wines by category (Red vs. White).
3. Category and Alcohol %:

- Examine the relationship between wine category and alcohol content.

4. Price of Wine and Alcohol %:
    - Investigate how alcohol content affects the price of wine.
5. Number of Wines by Rating:
    - Calculate the distribution of wines by their ratings.
6. Rating and Price of Wine:
    - Analyze the correlation between wine ratings and their prices.
7. Number of Wines by Country:
    - Determine the number of wines from each country in the dataset.
8. Price of Wine by Country:
    - Compare the average price of wines from different countries.

9. Wine Price by Rating and Country:
    - Explore the interaction between wine price, rating, and country of origin.
10. Actual Price vs. Predicted Price:
    - Develop a predictive model to estimate wine prices based on various featu res and compare predicted prices to actual prices.

Methodology:

1. Data Cleaning and Preprocessing:
    - Handle missing values and inconsistent data.

- Convert categorical variables to numerical values where necessary.

2. Exploratory Data Analysis (EDA):
   - Use statistical and visualization techniques to explore the data and identify patterns.
3. Data Visualization:
   - Create visualizations using libraries like Matplotlib and Seaborn to represe nt the findings.
4. Statistical Analysis:
   - Perform statistical tests to validate hypotheses and insights.
5. Predictive Modeling:

- Use machine learning algorithms (e.g., Linear Regression, Random Forest) to predict wine prices based on features such as alcohol content, rating, and country.

Tools and Technologies:

- Programming Language: Python
- Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- Visualization Tools: Tableau, Power BI

Expected Outcomes:

- Insightful visualizations and statistical summaries of the wine dataset.
- Understanding of how different attributes affect wine quality and pricing.
- Predictive model to estimate wine prices, providing actionable insights for wine enthusiasts and producers.

Timeline:

- Week 1: Data Collection and Cleaning
- Week 2: Exploratory Data Analysis and Visualization
- Week 3: Statistical Analysis and Hypothesis Testing
- Week 4: Predictive Modeling and Validation
- Week 5: Report Generation and Presentation