Project Report: Alcohol Consumption Analysis

1. Introduction

This project aims to analyze the alcohol consumption patterns from the provided consumption.csv dataset. The dataset comprises records detailing various bars, alcohol types, brand names, and their corresponding opening balances, purchases, consumption, and closing balances in milliliters. The primary objective is to extract meaningful insights into consumption trends, identify popular alcohol types and brands, assess bar performance, and pinpoint peak consumption times.

2. Data Overview

The consumption.csv dataset contains 6,575 entries across 8 distinct columns. A breakdown of the columns and their initial data types is as follows:

Column Name	Initial Data Type
Date Time Served	Object
Bar Name	Object
Alcohol Type	Object
Brand Name	Object
Opening Balance (ml)	Float64
Purchase (ml)	Float64
Consumed (ml)	Float64

Closing Balance (ml)	Float64

The dataset initially consumed approximately 411.1 KB of memory.

3. Exploratory Data Analysis (EDA)

3.1 Missing Values

A crucial initial step in data quality assessment is checking for missing values. Our comprehensive analysis revealed **no missing values** across any of the columns, ensuring a clean and robust dataset suitable for direct analysis without imputation.

3.2 Descriptive Statistics

Descriptive statistics for the numerical columns offer a preliminary understanding of the data's central tendency, dispersion, and shape:

Metric	Opening Balance (ml)	Purchase (ml)	Consumed (ml)	Closing Balance (ml)
count	6575	6575	6575	6575
mean	2468.40	315.84	299.42	2484.81
std	2284.55	582.12	191.90	2302.36
min	0.00	0.00	0.00	0.00
25%	619.14	0.00	156.64	611.00
50%	1848.44	0.00	300.39	1849.84

75%	3853.02	526.35	450.87	3906.05
max	11862.50	1999.84	1180.58	11862.50

- The Consumed (ml) column exhibits a mean consumption of approximately 299.42 ml per record, with a standard deviation of 191.9 ml, indicating a notable degree of variability in individual consumption instances.
- The Purchase (ml) column's descriptive statistics, particularly its minimum (0.00) and 25th percentile (0.00), suggest that a significant number of records do not involve new purchases, implying a focus on consumption from existing stock.

3.3 Unique Values in Categorical Columns

Understanding the distinct categories within the dataset's object type columns is essential:

- **Bar Name**: ["Smith's Bar", "Johnson's Bar", "Brown's Bar", "Taylor's Bar", "Anderson's Bar", "Thomas's Bar"]
  - There are **6 unique bars** represented in the dataset.
- **Alcohol Type**: ['Rum', 'Wine', 'Vodka', 'Beer', 'Whiskey']
  - The dataset includes **5 unique alcohol types**.
- **Brand Name**: ['Captain Morgan', 'Yellow Tail', 'Grey Goose', 'Coors', 'Jim Beam', 'Absolut', 'Jack Daniels', 'Sutter Home', 'Miller', 'Bacardi', 'Malibu', 'Jameson', 'Smirnoff', 'Budweiser', 'Heineken', 'Barefoot']
  - A total of **16 unique brands** are present in the consumption records.

4. Data Preprocessing and Feature Engineering

To facilitate comprehensive time-series analysis and enable time-based aggregations, the Date Time Served column underwent a critical transformation:

- It was converted from its initial 'Object' data type to a proper **datetime object**.
- Subsequently, dedicated Date and Time columns were extracted from the newly converted Date Time Served column, allowing for more granular analysis by date or specific time of day.

5. Key Findings and Analysis

5.1 Daily Consumption Trend

The daily consumption trend, as visualized in the line plot (refer to "Daily Consumption Trend"), showcases fluctuations over the observed period. While consumption on 2023-01-01 was recorded as zero, it notably increased in subsequent days, establishing an overall dynamic yet consistent pattern of consumption over time.

5.2 Total Consumption by Alcohol Type

The aggregate consumption for each alcohol type reveals distinct preferences:

Alcohol Type	Consumed (ml)
Vodka	411,760
Beer	399,498
Rum	393,267
Whiskey	384,754
Wine	379,402

Insight: **Vodka** stands out as the most consumed alcohol type, closely followed by Beer and Rum. Wine and Whiskey show slightly lower but comparable total consumption volumes. This trend suggests a robust market demand for Vodka and Beer within the observed establishments. (Refer to the "Total Consumption by Alcohol Type" bar plot for visual representation).

5.3 Total Consumption by Brand Name (Top 10)

An examination of consumption by brand highlights the top performers:

Brand Name	Consumed (ml)

Grey Goose	158,924
Jim Beam	139,702
Captain Morgan	138,036
Barefoot	132,313
Yellow Tail	129,844
Smirnoff	129,177
Bacardi	127,952
Malibu	127,280
Jack Daniels	125,051
Absolut	123,659

Insight: **Grey Goose** emerges as the undeniably top-consumed brand, indicative of its significant popularity and market presence. Other highly consumed brands include Jim Beam, Captain Morgan, and Barefoot. This breakdown provides specific brand-level performance insights, crucial for inventory and marketing decisions. (Refer to the "Top 10 Brands by Total Consumption" bar plot).

5.4 Total Consumption by Bar Name

An analysis of consumption across different bars provides an overview of individual bar performance:

Bar Name	Consumed (ml)
Johnson's Bar	344,184
Anderson's Bar	335,819
Smith's Bar	328,714
Brown's Bar	326,372
Thomas's Bar	318,416
Taylor's Bar	315,177

Insight: **Johnson's Bar** recorded the highest total consumption among all establishments, suggesting it might be the busiest or most frequented. Consumption figures are relatively close among the top bars, indicating a somewhat balanced distribution of activity, with Taylor's Bar registering the lowest total consumption. (Refer to the "Total Consumption by Bar Name" bar plot).

5.5 Total Consumption by Time of Day

Categorizing consumption by time of day reveals significant patterns:

Time of Day	Consumed (ml)

Morning	300,889
Afternoon	775,501
Evening	606,104
Night	286,187

Insight: The **Afternoon** period accounts for the highest consumption, followed by the **Evening**. Consumption is considerably lower during both Morning and Night hours. This pattern is vital for operational planning, indicating that bars experience their peak activity and demand during the afternoon and evening, which is critical for optimizing staffing levels and inventory allocation. (Refer to the "Total Consumption by Time of Day" bar plot).

6. Conclusion

This project successfully delivered a detailed and insightful analysis of alcohol consumption patterns derived from the consumption.csv dataset. Key conclusions drawn from this analysis include:

- The dataset is entirely clean and complete, requiring no data imputation.
- **Vodka**, **Beer**, and **Rum** consistently rank as the most popular alcohol types by consumption volume.
- **Grey Goose** stands out as the leading brand in terms of total consumption.
- **Johnson's Bar** demonstrated the highest overall consumption among all establishments.
- The **Afternoon** and **Evening** hours represent the peak consumption periods across the analyzed data.

These comprehensive insights provide invaluable information for bar management. They can be effectively utilized to optimize inventory management, strategically plan staffing, and tailor marketing strategies based on observed popular items, individual bar performance, and identified peak operational hours. Further analytical avenues could explore specific brand performance within each alcohol type, delve into purchase-to-consumption ratios, or conduct more granular daily trend analyses per bar.