

Tips for Data Analyst Report

This repository contains a collection of tips and best practices for creating effective data analyst reports. Whether you're a beginner or an experienced analyst, these tips will help you communicate your findings clearly and effectively.

1. Installation by create a python virtual environment

To set up a Python virtual environment for working with data analysis reports, follow these steps:

```
# Create a new virtual environment
python -m venv .venv
# Activate the virtual environment in windows using gitbash
source .venv/Scripts/activate

# create a requirements.txt file and add libraries inside
# Install required packages
pip install -r requirements.txt
```

or you can install the packages individually using pip:

```
pip install pandas numpy matplotlib seaborn plotly
pip install ipykernel openpyxl
```

2. Create a Jupyter Notebook

```
# create a ipynb file
mkdir scripts
cd scripts
touch 01_data_analysis.ipynb
```

2.1. Use of github copilot for data analytics report

Here is the prompt no. 1:

```
1. I have df in this notebook, which is tips data. Act as a data analyst, and write the code using python libraries pandas, numpy, seaborn, matplotlib and plot, to do the basic data analytics tasks such as:
```

1. Data composition report
2. Data distribution report
3. Data comparison report
4. Data relationship report

If you need to do the statistics please install stats models and scipy libraries using code cells in jupyter notebook.

If you want to create a plot and whenever you do that, please intepret the results as I would like to submit that in a report form Lets see what you have got, i am really looking forward to your code without bugs and best possible plots.

Complete report is here:

Restaurant Tips Data Analysis Report

Prepared by: Data Analytics Team
Date: January 2025
Dataset: Restaurant Tips Dataset (244 observations)

Executive Summary

This report presents a comprehensive analysis of restaurant tipping patterns based on a dataset of 244 dining transactions. The analysis examines data composition, distribution patterns, comparative behaviors across customer segments, and relationships between key variables. Key findings reveal strong correlations between bill amounts and tips, distinctive patterns across customer demographics, and actionable insights for restaurant operations.

Key Findings:

- Average tip percentage: 16.1%
 - Strong positive correlation between total bill and tip amount ($r=0.676$)
 - Weekend and dinner service generate higher revenue
 - Male customers and larger parties contribute to higher absolute tips
 - No significant data quality issues with complete records
-

1. Data Composition Analysis

Dataset Overview

The restaurant tips dataset contains **244 complete observations** across **7 variables**, representing a clean and well-structured dataset with no missing values.

Variable Structure

- **Numerical Variables (4):**
 - **total_bill**: Restaurant bill amount (\$3.07 - \$50.81)
 - **tip**: Tip amount (\$1.00 - \$10.00)
 - **size**: Party size (1-6 people)
 - **tip_percentage**: Calculated tip percentage (derived variable)
- **Categorical Variables (3):**

- **sex**: Customer gender (Male: 64%, Female: 36%)
- **smoker**: Smoking status (No: 76%, Yes: 24%)
- **day**: Day of week (Thur, Fri, Sat, Sun)
- **time**: Service time (Dinner: 68%, Lunch: 32%)

Data Quality Assessment

- **Completeness**: 100% - No missing values detected
- **Consistency**: All variables show expected ranges and formats
- **Accuracy**: No obvious outliers or data entry errors identified
- **Validity**: All categorical values fall within expected domains

[Insert Figure 1: Data Composition Summary Table]

2. Data Distribution Analysis

Numerical Variables Distribution

Total Bill Distribution:

- Mean: \$19.79, Median: \$17.80
- Right-skewed distribution indicating most bills are moderate with few high-value outliers
- Range: \$3.07 to \$50.81
- 75% of bills are under \$24.13

Tip Amount Distribution:

- Mean: \$2.98, Median: \$2.90
- Similar right-skewed pattern following bill distribution
- Range: \$1.00 to \$10.00
- Strong relationship with bill amount

Tip Percentage Distribution:

- Mean: 16.1%, Median: 15.5%
- Approximately normal distribution
- Most customers tip between 12-20%
- Few extreme cases below 5% or above 25%

Party Size Distribution:

- Mode: 2 people (most common)
- Range: 1-6 people
- 75% of parties are 2-3 people
- Larger groups (4+) represent opportunity for higher revenue

[Insert Figure 2: Distribution Plots for Numerical Variables] [Insert Figure 3: Box Plots for Outlier Detection]

Categorical Variables Distribution

Customer Demographics:

- **Gender Split:** 64% Male, 36% Female
- **Smoking Status:** 76% Non-smokers, 24% Smokers

Service Patterns:

- **Day Distribution:** Weekend-heavy with Saturday and Sunday being busiest
- **Time Distribution:** 68% Dinner service, 32% Lunch service

[Insert Figure 4: Categorical Variables Distribution Charts]

3. Data Comparison Analysis

Tipping Behavior by Customer Segments

Gender Comparison:

- Male customers: Average tip \$3.09
- Female customers: Average tip \$2.83
- Difference primarily attributed to higher bill amounts rather than generosity
- No statistically significant difference in tip percentages

Smoking Status Comparison:

- Smokers: Average tip \$3.01
- Non-smokers: Average tip \$2.97
- No significant difference in tipping behavior (p-value > 0.05)

Day of Week Analysis:

- **Highest Tips:** Sunday (\$3.26 average)
- **Lowest Tips:** Thursday (\$2.77 average)
- Weekend days show consistently higher absolute tips
- Statistical significance confirmed (ANOVA p-value < 0.05)

Service Time Comparison:

- **Dinner:** Average tip \$3.10
- **Lunch:** Average tip \$2.73
- Dinner customers tip 13.5% more on average
- Correlation with higher dinner bill amounts

[Insert Figure 5: Tip Comparison Across Categories] [Insert Figure 6: Tip Percentage Comparison Charts]

Statistical Significance Testing

- Gender differences: Not statistically significant (p > 0.05)
- Smoking status: No significant difference (p > 0.05)
- Day of week: Statistically significant variation (p < 0.05)
- Service time: Significant difference (p < 0.05)

4. Data Relationship Analysis

Correlation Analysis

Strong Positive Relationships:

- **Total Bill ↔ Tip Amount:** $r = 0.676$ (Strong)
 - For every \$10 increase in bill, tip increases by approximately \$1.83
 - Highly significant relationship ($p < 0.001$)
- **Party Size ↔ Total Bill:** $r = 0.598$ (Moderate-Strong)
 - Larger parties generate proportionally higher bills
 - Each additional person adds approximately \$7.44 to the bill

Interesting Inverse Relationships:

- **Total Bill ↔ Tip Percentage:** $r = -0.203$ (Weak Negative)
 - Higher bills receive slightly lower tip percentages
 - Customers may tip a flat amount rather than strict percentage

Moderate Relationships:

- **Party Size ↔ Tip Amount:** $r = 0.489$ (Moderate)
 - Larger groups tip more in absolute terms
 - Relationship mediated through bill amount

[Insert Figure 7: Correlation Heatmap] [Insert Figure 8: Scatter Plots with Regression Lines]

Advanced Relationship Patterns

Multi-variable Analysis:

- Dinner customers show stronger bill-to-tip correlation than lunch
- Male customers at dinner generate highest absolute tips
- Weekend parties of 3+ show premium tipping behavior
- Smoking status shows no interaction effects with other variables

Categorical Variable Associations:

- Gender and smoking status: Independent ($\chi^2 p > 0.05$)
- Day and time: Significant association ($\chi^2 p < 0.001$)
- Time preferences vary by day of week

[Insert Figure 9: Advanced Relationship Visualizations]

5. Business Insights and Recommendations

Revenue Optimization

1. **Focus on Dinner Service:** 68% of business with higher per-customer value
2. **Weekend Marketing:** Saturday/Sunday show peak performance
3. **Party Size Strategy:** Target groups of 3+ for maximum revenue impact
4. **Bill Scaling:** Understand that tip percentages may decrease on high-value bills

Operational Insights

1. **Staffing Patterns:** Align staff levels with dinner/weekend peaks
2. **Customer Segmentation:** Male dinner customers and larger parties are high-value segments
3. **Service Quality:** Maintain consistent service regardless of bill size
4. **Capacity Planning:** Weekend dinner combinations require maximum capacity

Strategic Recommendations

1. **Loyalty Programs:** Target frequent dinner customers and larger parties
 2. **Menu Pricing:** Consider tip percentage decline on higher-value items
 3. **Service Training:** Focus on consistent percentage-based tipping education
 4. **Marketing Timing:** Promote weekend dinner experiences for maximum ROI
-

6. Limitations and Future Analysis

Current Limitations

- Single restaurant/chain data (generalizability questions)
- Limited time period (seasonal variations unknown)
- No customer satisfaction or service quality metrics
- Missing payment method information

Recommended Future Analysis

1. **Seasonal Patterns:** Year-round data collection
 2. **Service Quality Metrics:** Customer satisfaction correlation with tips
 3. **Payment Methods:** Cash vs. card tipping behavior analysis
 4. **Competitive Analysis:** Industry benchmarking
 5. **Economic Factors:** Impact of economic conditions on tipping
-

7. Conclusions

The restaurant tips dataset analysis reveals clear patterns in customer behavior and tipping practices. The strong correlation between bill amounts and tips provides a reliable foundation for revenue forecasting, while the identification of high-value customer segments (dinner customers, weekend visitors, larger parties) offers actionable insights for business optimization.

Key strategic takeaways include the importance of dinner service optimization, weekend capacity management, and understanding the nuanced relationship between bill size and tip percentages. The analysis supports data-driven decision making for restaurant operations, marketing strategies, and service delivery improvements.

The clean, complete nature of the dataset and the statistical significance of key relationships provide confidence in these findings and their applicability to restaurant business strategy.

Report Prepared Using:

- Python 3.x with pandas, numpy, matplotlib, seaborn
- Statistical analysis with scipy and statsmodels
- Data visualization and correlation analysis
- Hypothesis testing and significance evaluation

Note: All figures referenced in this report are generated from the accompanying Jupyter notebook analysis and should be inserted at the indicated locations for the complete PDF version.

These changes are made with AI (Github Copilot).

 **Author: Dr. Muhammad Aamamr Tufail**

 GITHUB

PROFILE

 KAGGLE

PROFILE

 LINKEDIN

PROFILE

 YOUTUBE

PROFILE

 FACEBOOK

PROFILE

 TIKTOK

PROFILE

 TWITTER

PROFILE

 INSTAGRAM

PROFILE

 EMAIL

CONTACT ME