

Trip Expense Data Analysis Report

Executive Summary

This report presents a comprehensive analysis of 1,000 trip expense records, examining the relationships between trip duration, miles traveled, receipt amounts, and expected outputs (total trip costs). The analysis reveals several key patterns and insights that can inform trip planning and expense management decisions.

Dataset Overview

The dataset contains 1,000 trip records with the following characteristics:

- **Trip Duration:** Ranges from 1 to 14 days (average: 7.04 days)
- **Miles Traveled:** Ranges from 13 to 1,061 miles (average: 592.6 miles)
- **Total Receipts Amount:** Ranges from \$1.42 to \$2,066.62 (average: \$611.85)
- **Expected Output (Total Cost):** Ranges from \$117.24 to \$2,337.73 (average: \$1,349.11)

Key Findings

1. Trip Duration Impact on Costs

The analysis reveals a strong positive correlation between trip duration and total costs, but with diminishing returns on a per-day basis:

- **Total Cost Growth:** Longer trips have higher total costs, with the relationship being roughly linear but with some variation
- **Cost Per Day Efficiency:** Shorter trips (1-2 days) have significantly higher cost per day (\$873.55 for 1-day trips vs \$121.93 for 14-day trips)
- **Optimal Duration:** Trips of 5+ days show more cost-efficient daily rates, suggesting economies of scale for longer trips

2. Miles Traveled Analysis

Distance traveled shows interesting patterns across different trip durations:

- **Moderate Correlation:** Miles traveled has a moderate positive correlation (0.43) with total costs

- **Variable Efficiency:** Cost per mile varies significantly, ranging from under \$1 to over \$300 per mile
- **Duration Independence:** Average miles traveled doesn't consistently increase with trip duration, suggesting different trip purposes

3. Receipt Patterns

Receipt amounts as a percentage of total trip costs reveal important insights:

- **Average Coverage:** Receipts typically represent 82-110% of total trip costs
- **Variation by Duration:** 1-day trips show the highest receipt percentages (109.96%), while longer trips show more moderate percentages
- **Distribution:** Most trips have receipt percentages between 50-150%, with some outliers

4. Cost Efficiency Trends

Several efficiency patterns emerge from the data:

- **Economies of Scale:** Longer trips generally offer better cost per day efficiency
- **Variable Mile Costs:** Cost per mile shows high variability, suggesting different trip types or purposes
- **Sweet Spot:** Trips of 5-10 days appear to offer the best balance of total value and daily efficiency

Detailed Analysis by Visualization

Distribution Analysis

Expected Output Distribution: The total trip costs follow a roughly normal distribution with a slight right skew, centered around \$1,400-\$1,500. This suggests most trips fall within a predictable cost range, with some high-cost outliers.

Cost Per Mile Distribution: Shows a highly right-skewed distribution with most trips costing \$1-\$5 per mile, but with significant outliers reaching \$300+ per mile. This suggests different trip types (business vs leisure, urban vs rural, etc.).

Cost Per Day Distribution: Exhibits a strong right skew with most trips costing \$100-\$300 per day, but with a long tail extending to over \$1,400 per day for very short trips.

Correlation Analysis

The correlation matrix reveals several important relationships:

- **Trip Duration ↔ Expected Output:** Moderate positive correlation (0.51)
- **Miles Traveled ↔ Expected Output:** Moderate positive correlation (0.43)
- **Receipts ↔ Expected Output:** Strong positive correlation (0.70)
- **Duration ↔ Miles:** Weak correlation (0.05), indicating trip length and distance are largely independent

Trend Analysis by Trip Duration

1-2 Day Trips: - Highest cost per day (\$873.55 and \$523.12 respectively) - High receipt percentages (109.96% and 95.57%) - Likely business trips or short getaways with premium pricing

3-5 Day Trips: - Moderate cost efficiency - Balanced receipt percentages (82-84%) - Typical vacation or extended business trip duration

6-10 Day Trips: - Improved cost efficiency (\$227.75 to \$149.61 per day) - Lower receipt percentages (72-86%) - Extended vacations or business assignments

11-14 Day Trips: - Best cost efficiency (\$121.93 to \$145.51 per day) - Moderate receipt percentages (69-85%) - Long-term assignments or extended travel

Business Implications

For Trip Planning

1. **Budget Optimization:** Trips of 5+ days offer significantly better value per day
2. **Cost Predictability:** Most trips fall within predictable cost ranges, enabling better budgeting
3. **Receipt Management:** Expect receipts to cover 80-110% of total costs for most trips

For Expense Management

1. **Anomaly Detection:** Trips with cost per mile >\$50 or cost per day >\$500 warrant review
2. **Duration Strategy:** Consider extending short trips to 3+ days for better cost efficiency

3. **Receipt Tracking:** 1-day trips may have receipt percentages >100% due to setup costs

For Policy Development

1. **Per Diem Rates:** Should vary significantly by trip duration, with higher rates for shorter trips
2. **Approval Thresholds:** Different thresholds needed for different trip durations
3. **Efficiency Metrics:** Cost per day is more meaningful than total cost for evaluation

Visualization Insights

Comprehensive Analysis Chart

The comprehensive analysis chart provides a multi-faceted view of the trip data:

1. **Distribution Histograms:** Show the spread and central tendencies of key metrics
2. **Scatter Plots:** Reveal relationships between variables and identify clusters
3. **Box Plots:** Highlight the median, quartiles, and outliers by trip duration
4. **Correlation Matrix:** Quantifies the strength of relationships between variables

Key insights from this visualization include: - The expected output (total cost) distribution peaks around \$1,500 - Trip duration and expected output show a clear positive relationship - Miles traveled and expected output show a moderate positive relationship - Receipt amounts and expected output show a strong positive relationship - Cost per mile is highly concentrated at the lower end of the scale - Cost per day decreases significantly as trip duration increases

Detailed Scatter Plots

The detailed scatter plots use color coding to reveal multi-dimensional relationships:

1. **Duration vs Output (Color: Miles):** Shows how miles traveled influences the duration-cost relationship
2. **Miles vs Output (Color: Duration):** Reveals how trip duration affects the miles-cost relationship
3. **Receipts vs Output (Color: Duration):** Demonstrates how trip duration impacts the receipts-cost relationship
4. **Cost per Mile vs Cost per Day (Color: Duration):** Highlights efficiency metrics across different trip durations

Key insights from these visualizations include: - Longer trips (yellow points) tend to have higher costs regardless of miles traveled - Higher mileage correlates with higher costs, but with significant variation - Receipt amounts strongly predict total costs, especially for longer trips - Short trips (purple points) show extreme variability in cost per mile and cost per day

Trend Analysis Charts

The trend analysis charts track key metrics across different trip durations:

1. **Average Expected Output:** Shows a generally increasing trend with trip duration
2. **Average Cost per Day:** Reveals a dramatic decrease as trip duration increases
3. **Average Cost per Mile:** Shows significant variability across trip durations
4. **Average Receipts Percentage:** Demonstrates how receipt coverage varies by trip duration

Key insights from these trend analyses include: - Total cost increases with trip duration, but not proportionally - Cost per day efficiency improves dramatically for longer trips - Cost per mile shows irregular patterns, suggesting different trip purposes - Receipt percentages are highest for very short trips and show variability for longer trips

Conclusion and Recommendations

Summary of Findings

This analysis reveals that trip expenses follow predictable patterns based on duration, distance, and receipt amounts. The data suggests significant economies of scale for longer trips, with cost efficiency improving dramatically as trip duration increases. The relationship between miles traveled and total cost is positive but moderate, indicating that other factors beyond distance significantly influence trip expenses.

Recommendations for Further Analysis

1. **Segmentation Analysis:** Further segment the data by trip purpose, location, or traveler type
2. **Predictive Modeling:** Develop a predictive model for trip costs based on duration, distance, and receipt amounts
3. **Anomaly Detection:** Implement an anomaly detection system to identify unusual expense patterns
4. **Temporal Analysis:** If timestamp data becomes available, analyze seasonal or time-based patterns

5. **Comparative Benchmarking:** Compare these patterns against industry standards or historical data

Next Steps

1. **Interactive Dashboard:** Develop an interactive dashboard for real-time expense analysis
2. **Policy Refinement:** Use these insights to refine travel policies and approval thresholds
3. **Budget Optimization:** Implement duration-based budgeting for more accurate financial planning
4. **Traveler Education:** Share efficiency insights with travelers to encourage cost-effective planning

This analysis provides a solid foundation for understanding trip expense patterns and can serve as a basis for more sophisticated expense management strategies and policies.

Report generated on June 7, 2025