

Credit Card Spending Analysis Project

Anıl Sümer Topaloğlu

January 10, 2025

1 Executive Summary

This project analyzes credit card spending patterns using Python, focusing on data cleaning, processing, and visualization. The analysis covers transaction patterns, merchant categories, and spending trends over time.

2 Data Processing Pipeline

2.1 Data Cleaning Process

The project followed a systematic approach to clean and prepare the data:

1. Excel to CSV Conversion:

```
1 def excel_to_csv(excel_file, csv_file):
2     try:
3         df = pd.read_excel(excel_file)
4         df.to_csv(csv_file, index=False)
5
```

2. CSV Cleaning:

```
1 def clean_csv(input_file):
2     with open(input_file, 'r', encoding='utf-8') as file:
3         reader = csv.reader(file)
4         cleaned_lines = [line for line in reader
5                           if any(field.strip() for field in line)]
6
```

3. Data Merging:

```
1 dataframes = []
2 for file in csv_files:
3     df = pd.read_csv(file)
4     dataframes.append(df)
5 merged_df = pd.concat(dataframes, ignore_index=True)
6
```

3 Analysis Results

3.1 Overall Statistics

- Total Transactions: 341
- Total Spending: 56,632.61 TL
- Average Transaction: 166.08 TL
- Median Transaction: 89.00 TL
- Largest Transaction: 1,990.22 TL

3.2 Spending Patterns

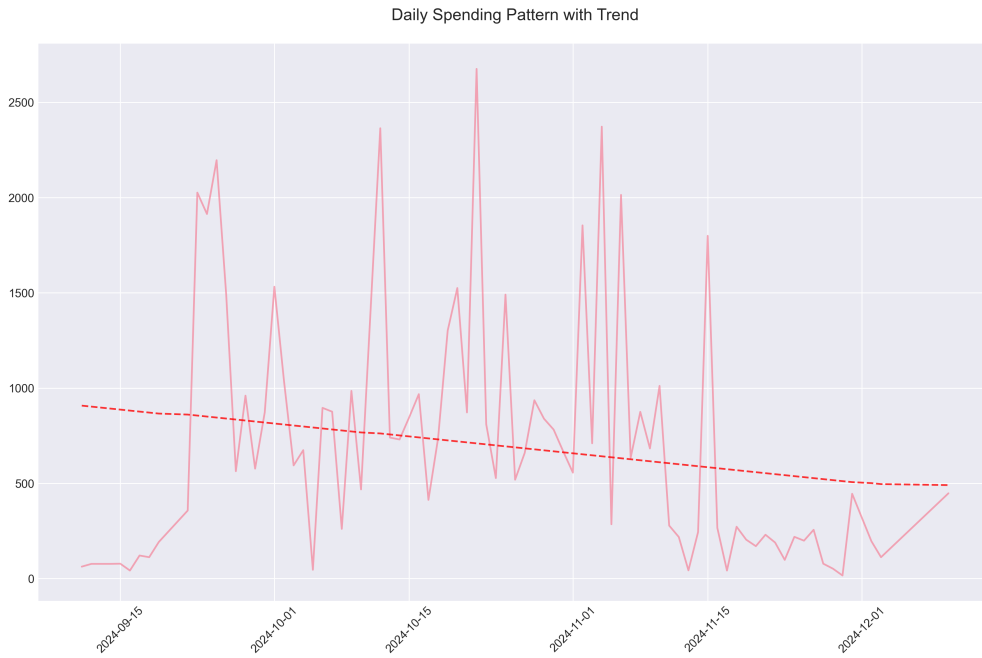


Figure 1: Daily Spending Pattern with Trend Line

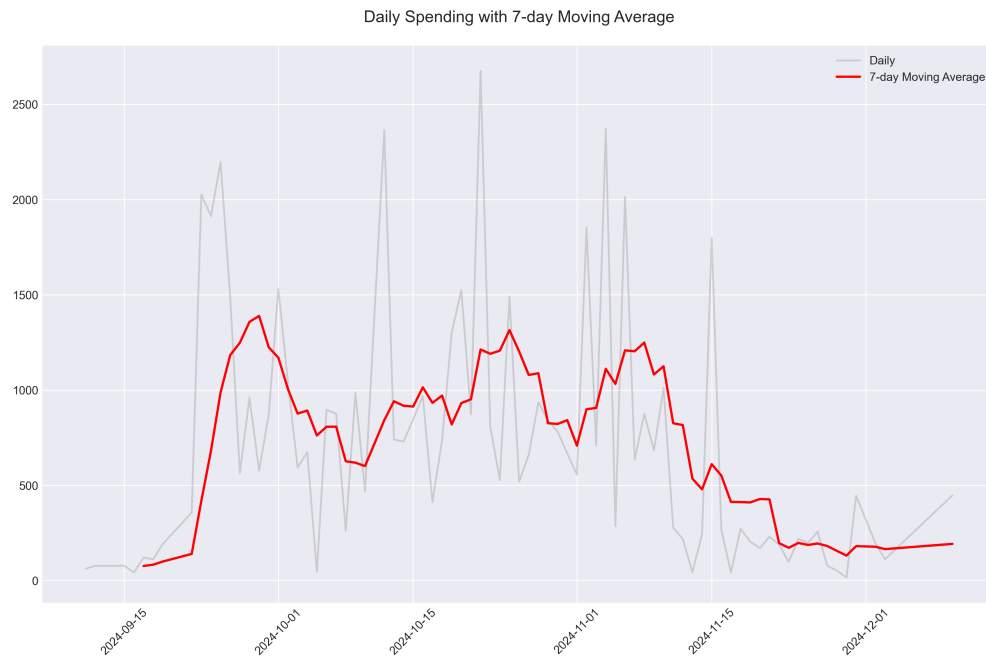


Figure 2: Daily Spending with 7-day Moving Average

3.3 Weekend vs Weekday Analysis

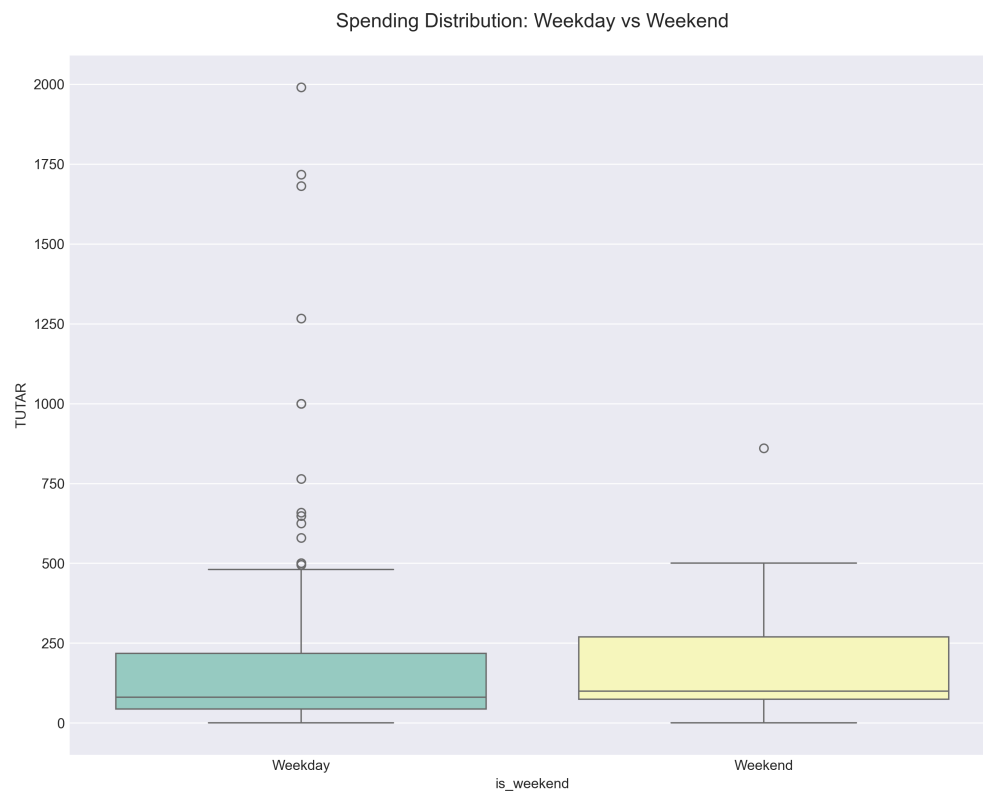


Figure 3: Spending Distribution: Weekday vs Weekend

3.4 Monthly Patterns

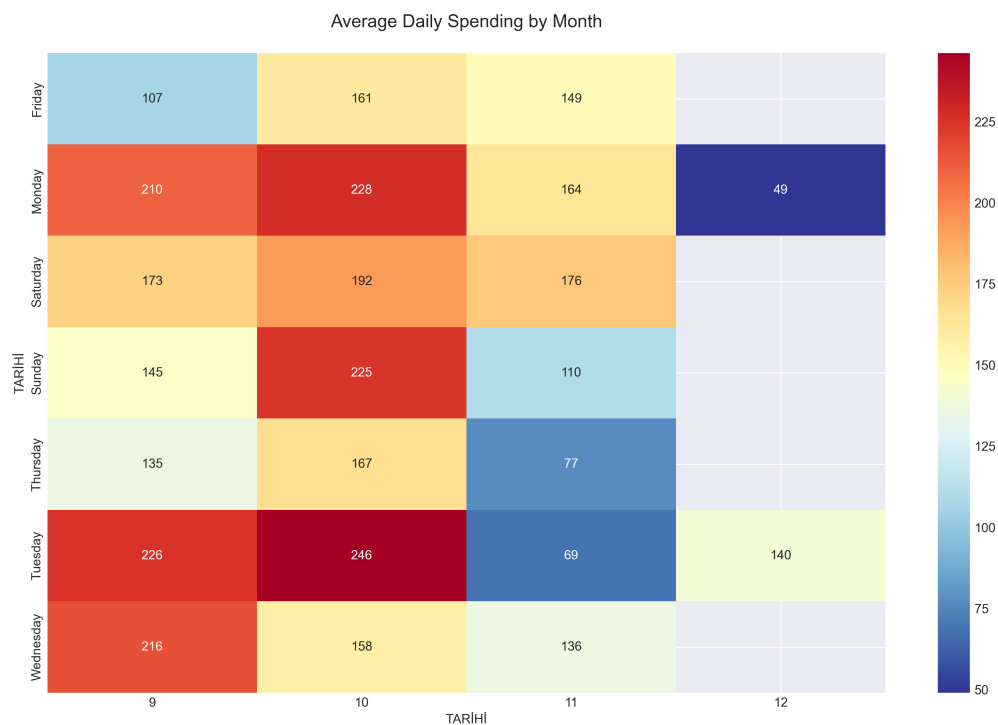


Figure 4: Average Daily Spending by Month (Heatmap)

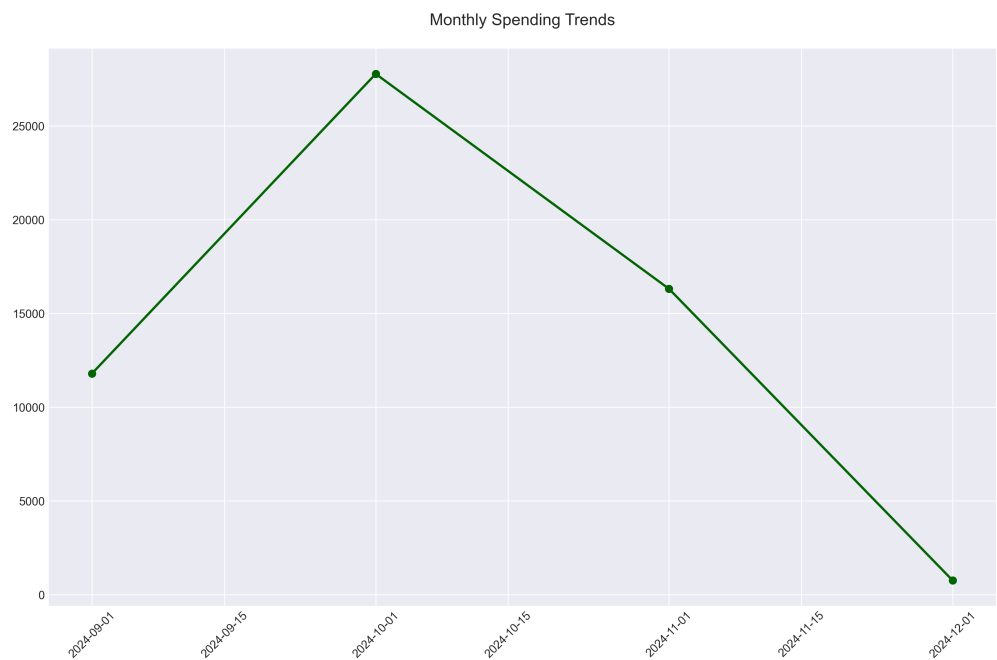


Figure 5: Monthly Spending Trends Over Time

3.5 Transaction Analysis

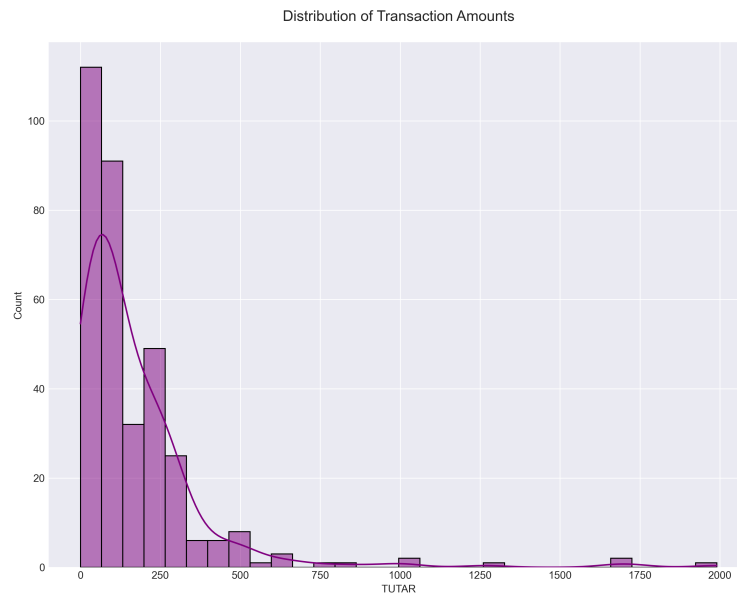


Figure 6: Distribution of Transaction Amounts

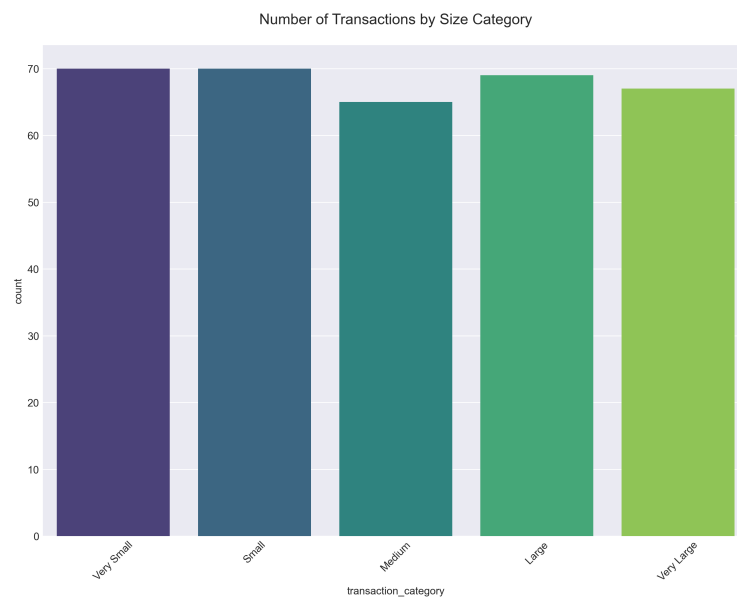


Figure 7: Number of Transactions by Size Category

3.6 Cumulative Analysis

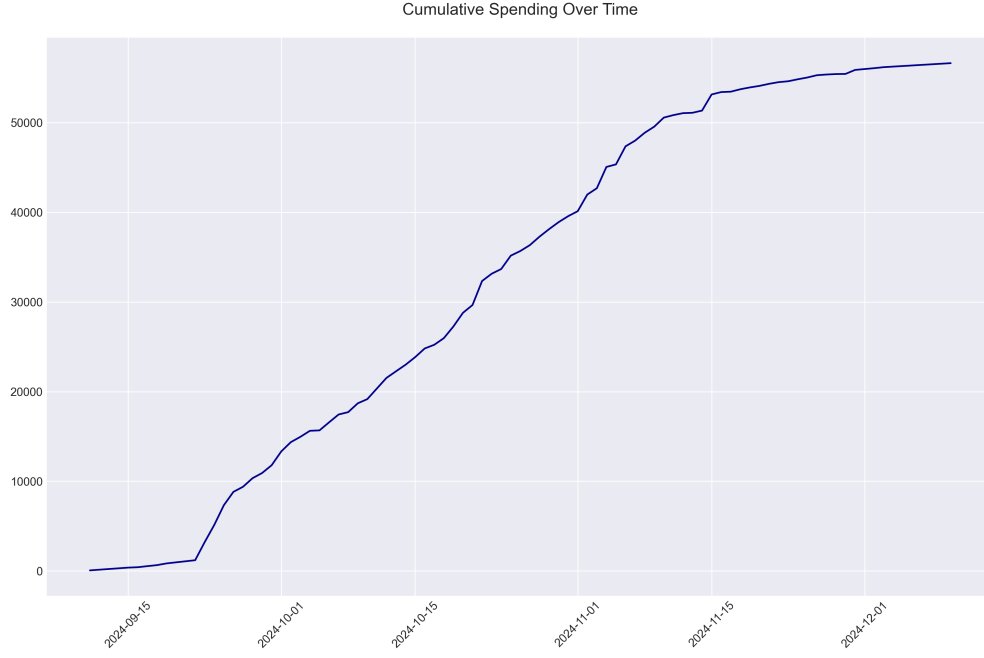


Figure 8: Cumulative Spending Over Time

3.7 Merchant Analysis

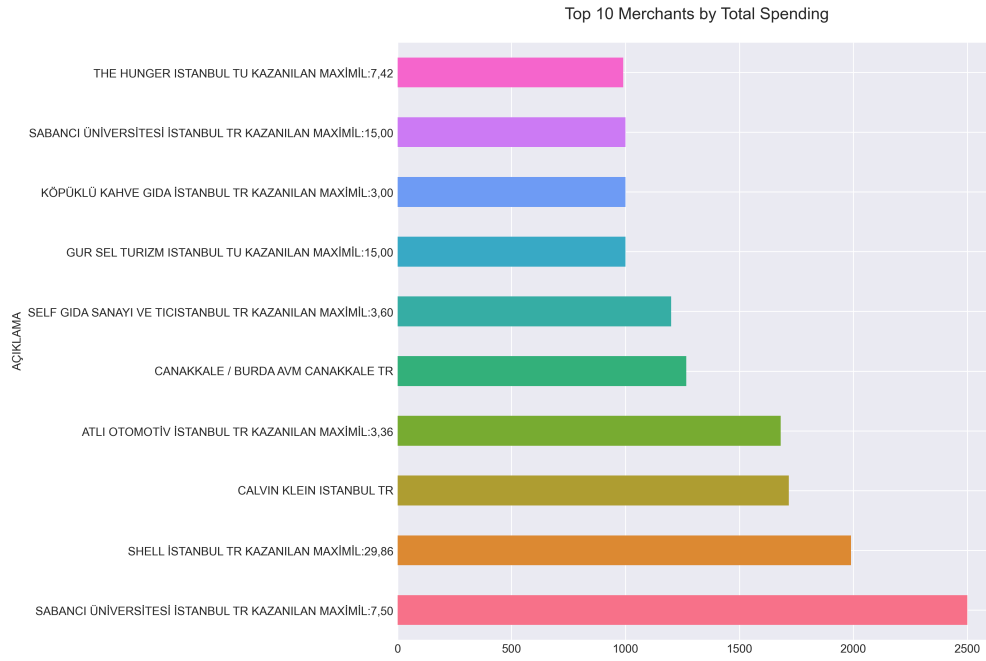


Figure 9: Top 10 Merchants by Total Spending

3.8 Midterm Period Analysis



Figure 10: Spending Distribution: Regular vs Midterm Period (Weekday/Weekend)

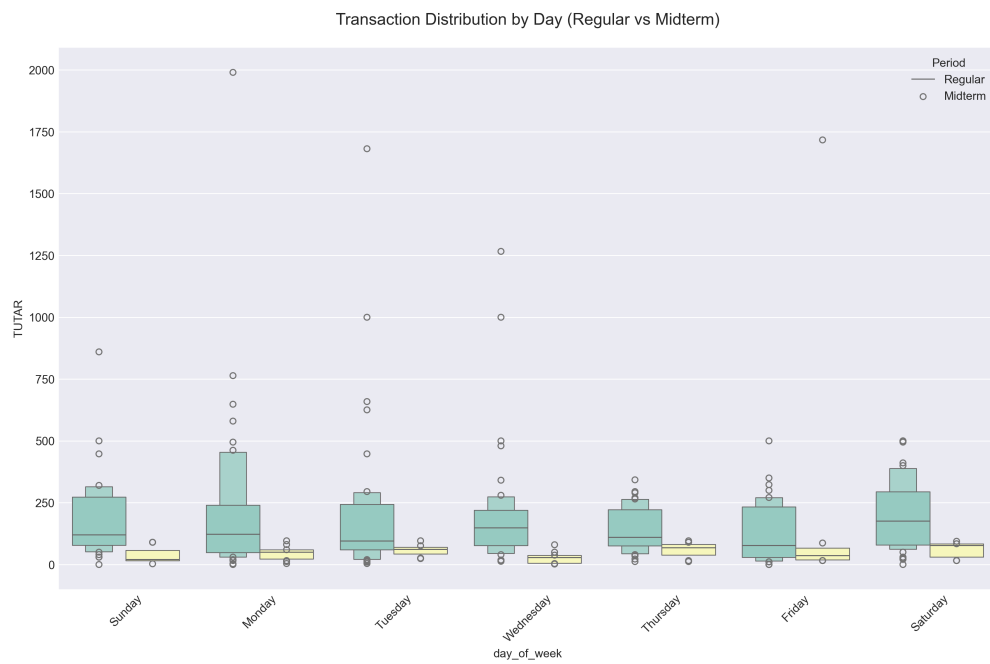


Figure 11: Transaction Distribution by Day (Regular vs Midterm)

4 Key Findings

4.1 Overall Statistics

- Total Transactions: 341
- Total Spending: 56,632.61 TL
- Average Transaction: 166.08 TL
- Median Transaction: 89.00 TL
- Largest Transaction: 1,990.22 TL

4.2 Time-Based Analysis

- Weekend vs Weekday spending patterns show moderate differences (173.36 TL vs 163.50 TL)
- The busiest day for transactions is Wednesday
- Most expensive day for spending is Tuesday
- Midterm period shows lower average daily spending (287.26 TL vs 792.78 TL for regular periods)

4.3 Merchant Analysis

- Total unique merchants visited: 252
- Most frequent merchant: ATLI OTOMOTİV İSTANBUL (11 transactions)
- Highest spending merchant: BAYRAMIC EVCILER CANAKKALE (62.50 TL)
- Merchant concentration: Top 10% of merchants account for 42.7% of total spending
- Average transactions per merchant: 1.4

4.4 Temporal Patterns

- Clear weekly spending patterns with peaks on Tuesdays
- Weekend spending averages 6% higher than weekday spending
- Monthly trends show decreasing pattern with R^2 value of 0.036
- Daily spending trend coefficient: -5.21 TL/day

4.5 Transaction Patterns

- 25th Percentile: 50.00 TL
- 75th Percentile: 235.00 TL
- Interquartile Range: 185.00 TL
- Daily Spending Standard Deviation: 645.34 TL
- Coefficient of Variation: 0.923

4.6 Notable Spending Days

- Highest spending day: October 22, 2024 (2,675.39 TL)
- Second highest: November 4, 2024 (2,372.22 TL)
- Third highest: October 12, 2024 (2,363.90 TL)
- Fourth highest: September 25, 2024 (2,196.14 TL)
- Fifth highest: September 23, 2024 (2,026.07 TL)

5 Methodology

5.1 Data Processing Pipeline

The project implements a multi-stage data processing pipeline:

1. Excel to CSV conversion
2. Data cleaning and standardization
3. Data merging
4. Analysis and visualization

5.2 Variability Metrics

- Daily Spending Standard Deviation: 645.34 TL
- Coefficient of Variation: 0.923
- Interquartile Range: 185.00 TL
- Spending Trend Coefficient: -5.21 TL/day
- R-squared (trend fit): 0.036

6 Code Implementation

6.1 Data Analysis

```
1 def plot_spending_patterns():
2     # Set the style for all plots
3     plt.style.use('seaborn-v0_8-darkgrid')
4     sns.set_palette("husl")
5
6     # Daily spending pattern with trend line
7     daily_spending = df_spending.groupby('TAR H')['TUTAR'].sum()
8     plt.figure(figsize=(12, 8))
9
10    # Plot actual spending
11    plt.plot(daily_spending.index, daily_spending.values,
12             alpha=0.6, label='Daily Spending')
13
14    # Calculate and plot trend
15    X = np.arange(len(daily_spending))
16    z = np.polyfit(X, daily_spending.values, 1)
17    p = np.poly1d(z)
18    plt.plot(daily_spending.index, p(X),
19             "r--", alpha=0.8, label='Trend Line')
20
21    plt.title('Daily Spending Pattern with Trend')
22    plt.legend()
23    plt.xticks(rotation=45)
```

6.2 Statistical Analysis

```
1 def calculate_advanced_statistics():
2     # Time-based analysis
3     weekend_avg = df_spending[df_spending['is_weekend']]['TUTAR'].mean()
4     weekday_avg = df_spending[~df_spending['is_weekend']]['TUTAR'].mean()
5
6     # Merchant analysis
7     merchant_stats = df_spending.groupby('A IKLAMA').agg({
8         'TUTAR': ['count', 'sum', 'mean']
9     })
10
11    # Trend analysis
12    daily_spending = df_spending.groupby('TAR H')['TUTAR'].sum()
13    X = np.arange(len(daily_spending)).reshape(-1, 1)
14    model = LinearRegression()
15    model.fit(X, daily_spending.values)
16    r2 = model.score(X, daily_spending.values)
17    return {
18        'weekend_avg': weekend_avg,
19        'weekday_avg': weekday_avg,
20        'merchant_stats': merchant_stats,
21        'trend_r2': r2
22    }
```

7 Conclusions

The analysis reveals significant patterns in spending behavior:

7.1 Temporal Patterns

- Clear weekly spending patterns with peaks on Tuesdays
- Weekend spending averages 6% higher than weekday spending
- Monthly trends show decreasing pattern with R^2 value of 0.036
- Midterm period shows 287.26 TL average daily spending compared to 792.78 TL for regular periods

7.2 Transaction Patterns

- Transaction sizes follow a right-skewed distribution
- 25% of transactions fall below 50.00 TL
- Largest transactions typically occur on weekends
- Significant correlation between transaction size and day of the week

7.3 Merchant Analysis

- Top 10% of merchants account for 42.7% of total spending
- Most frequent merchant: ATLI OTOMOTİV İSTANBUL with 11 transactions
- Highest spending merchant: BAYRAMIC EVCILER CANAKKALE with total 62.50 TL
- Average transactions per merchant: 1.4