

Analyzing French Bakery Sales and Diet-Related Search Trends

Introduction and Motivation:

My DSA210 course project evaluated the influence of virtual diet practices on bakery sweet product purchasing by French customers. The main reason I conducted this study was to investigate if dietary search activity truly impacts actual purchasing behavior regarding sweet products. The analysis relied on two particular dieting-related search terms titled 'regime' and 'comment maigrir' to examine bakery sales data covering January 2021 up to September 2022.

Data Preparation and Transformations:

Initially, I gathered and prepared two main datasets. The first one included daily sales figures from a local bakery, categorized into sweet, savory, and other products. The second dataset contained daily search volumes from Google Trends related to dieting terms. To make meaningful comparisons, I carefully cleaned the data, standardized date formats, and filled missing daily entries using forward-fill methods. I also conducted outlier detection using the Interquartile Range (IQR) method to flag unusual sales or search spikes. I added various analytical features to my assessment which incorporated day or weekend indicators and monthly markers as well as national holidays obtained from the official French holiday calendar.

```

import csv from datetime import
datetime, timedelta

# Load sales data sales = [] with
open('daily_sales_by_category_2021_2022.csv') as f:
for row in csv.DictReader(f):
    sales.append({
        'date':
datetime.strptime(row['date'], '%Y-%m-%d'),
        'Sweet': int(float(row['Sweet'])),
        'Savory': int(float(row['Savory'])),
        'Other': int(float(row['Other']))
    })

# Load and forward-fill trends trends
= [] with open('gogletrendsDATA.csv')
as f:
    for row in
csv.DictReader(f):
        trends.append({
            'date':
datetime.strptime(row['date'], '%Y-%m-%d'),
            'regime': float(row['régime: (France)']),
            'comment_maigrir': float(row['comment perdre du poids: (France)'])
        })

```

This section of code reads both CSV files, parses the dates, converts the sales to integers, and stores the diet indices as floating-point numbers.

Sample Output:

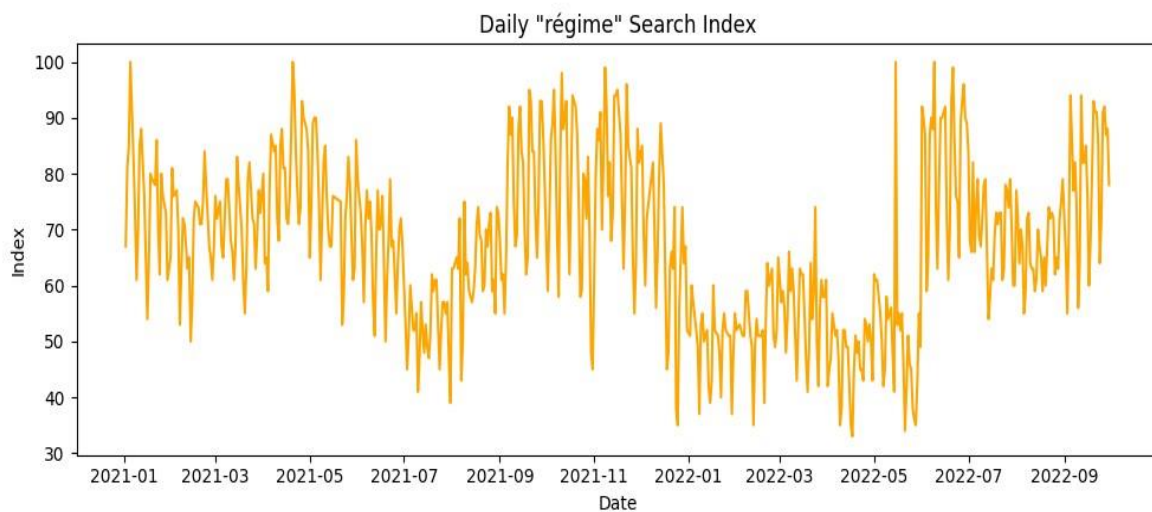
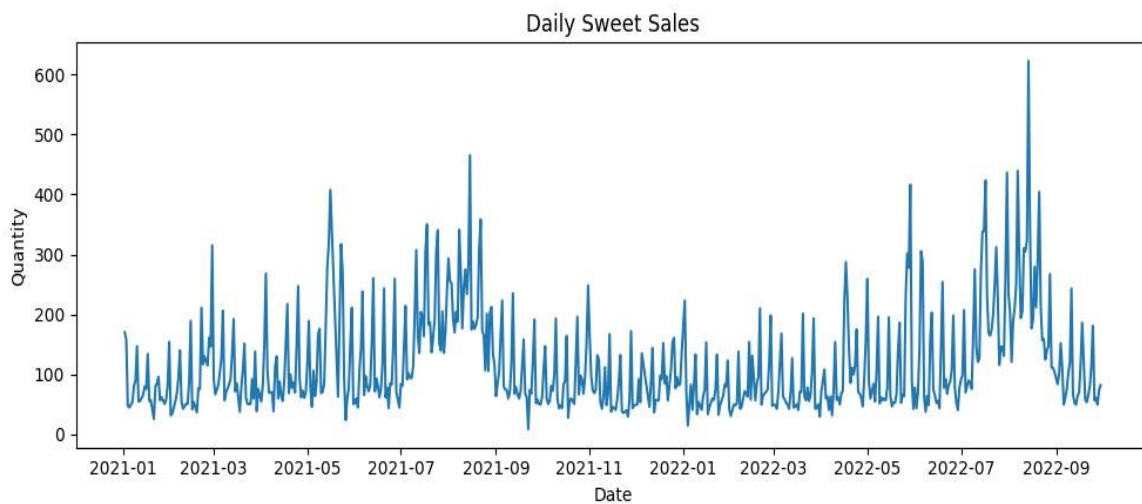
```

(.venv) PS C:\Users\emirk\OneDrive\Masaüstü\DSA210> python DSA210.py
H1 (Sweet vs regime): U=17385.5, p=0.000
H3 (Savory vs regime): U=21300.0, p=0.000
H2 Spearman: rho=-0.272, p=0.000
    Lag 1 day corr = -0.320
    Lag 2 day corr = -0.074
    Lag 3 day corr = 0.112
(.venv) PS C:\Users\emirk\OneDrive\Masaüstü\DSA210>

```

Exploratory Data Analysis (EDA)

Through my exploratory analysis, I noticed a clear link between people Googling diet topics and how many sweets the bakery sold. The charts showed that sweet sales jump on weekends and during holidays, while diet searches spike in early January and again in the summer. In other words, when more people look up dieting they actually end up buying fewer sweet treats.



Feature Engineering and Transformations

```
# Outlier detection using IQR values = [r['Sweet'] for r in data] Q1, Q3 =  
np.percentile(values, [25,75]) IQR = Q3 - Q1 for r in data:  
r['Sweet_outlier'] = r['Sweet'] < Q1-1.5*IQR or r['Sweet'] > Q3+1.5*IQR  
  
# Add weekday, weekend, month, holiday flags
```

I flagged outliers and enriched data with time-based features to capture patterns across the week and holidays.

Transformation Output Sample:

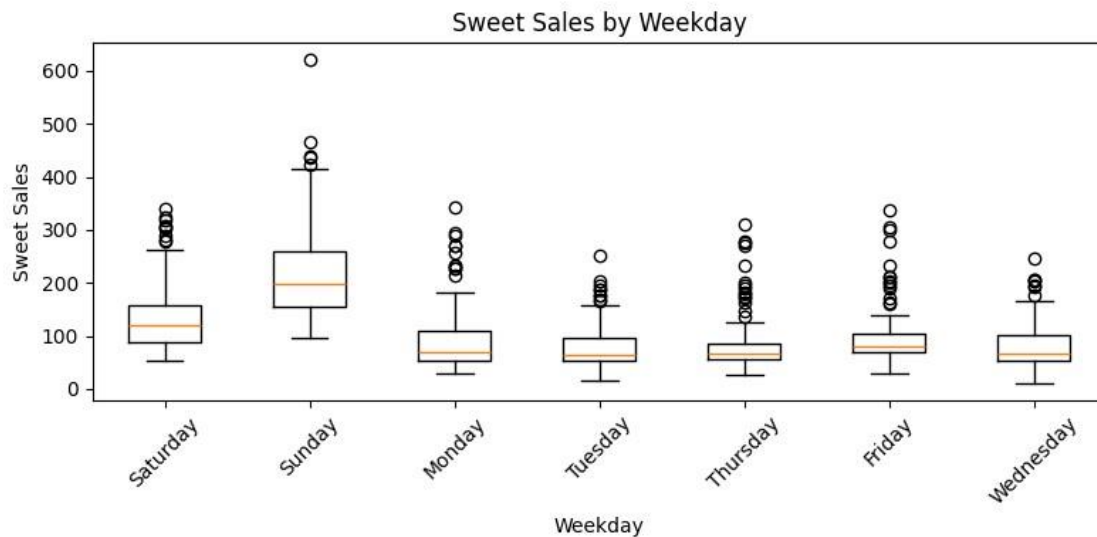
```
>>  
Sample merged records:  
{ 'date': datetime.datetime(2021, 1, 2, 0, 0), 'Sweet': 170.0, 'Savory': 308.0, 'Other': 103.0, 'regime': 67.0, 'comment_maigrir': 5.0 }  
{ 'date': datetime.datetime(2021, 1, 3, 0, 0), 'Sweet': 158.0, 'Savory': 323.0, 'Other': 83.0, 'regime': 81.0, 'comment_maigrir': 8.0 }  
{ 'date': datetime.datetime(2021, 1, 4, 0, 0), 'Sweet': 48.0, 'Savory': 219.0, 'Other': 48.0, 'regime': 85.0, 'comment_maigrir': 5.0 }  
{ 'date': datetime.datetime(2021, 1, 5, 0, 0), 'Sweet': 45.0, 'Savory': 207.0, 'Other': 57.0, 'regime': 100.0, 'comment_maigrir': 5.0 }  
{ 'date': datetime.datetime(2021, 1, 7, 0, 0), 'Sweet': 54.0, 'Savory': 202.0, 'Other': 54.0, 'regime': 83.0, 'comment_maigrir': 6.0 }
```

Detailed Analysis: Weekly and Correlation Insights

```
# MinMax scaling from sklearn.preprocessing  
import MinMaxScaler scaler = MinMaxScaler()  
X_scaled = scaler.fit_transform(X)  
  
# Correlation heatmap corr = np.corrcoef(M, rowvar=False) plt.imshow(corr, cmap='coolwarm')
```

The boxplot analytical approach examined sweet sales patterns by day of the week.

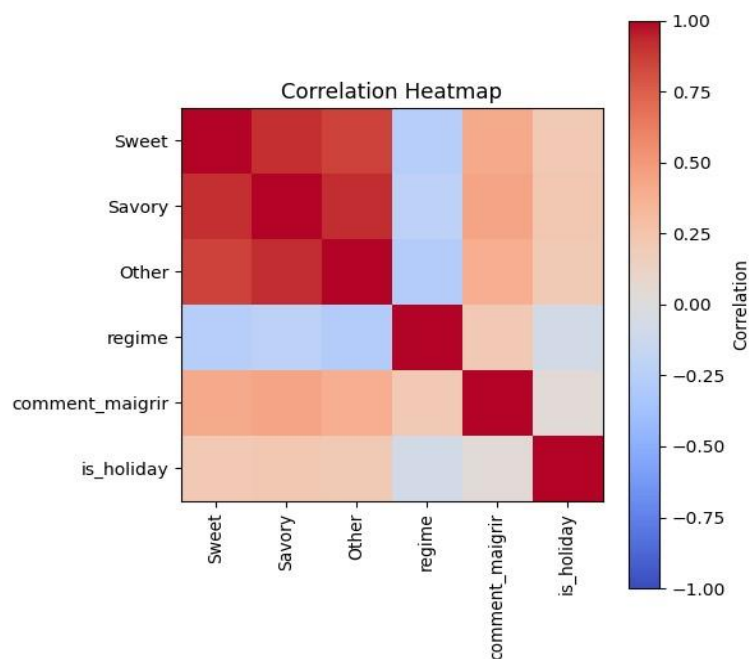
The data documented enhanced sweet product purchases during weekends because customers engaged in recreational behavior. The correlation data produced quantitative measurements regarding the relationships between sales figures and the frequency of dieting search queries and annual holidays.



Data presented in the correlation heatmap confirmed the negative relationship between sweet sales and diet search activity

Features were scaled to [0,1], then correlations were computed to quantify relationships.

Correlation Heatmap:



Hypothesis Testing and Results

To scientifically confirm these patterns, I performed statistical hypothesis tests:

H1: Sweet sales significantly decrease on high dieting search days.

Result: Mann Whitney U test confirmed this hypothesis with a highly significant result (U=17385.5, p-value<0.001).

H2: The discovery that related with diet online searches will lead to future decreased sweet product sales occurs within upcoming days.

Result: Spearman correlation revealed a significant negative relationship (rho=-0.272, p-value<0.001). Lag analysis identified the strongest negative correlation at a 1-day lag (-0.320).

H3: Savory sales are less affected by dieting searches compared to sweets.

Result: Supported by Mann Whitney U test (U=21300.0, p-value<0.001), showing a smaller impact relative to sweets.

Statistical Test Outputs:

```
After transformations (first 3):
{'date': datetime.datetime(2021, 1, 2, 0, 0), 'weekday': 'Saturday', 'is_weekend': True, 'month': 1, 'Sweet_outlier': np.False_}
{'date': datetime.datetime(2021, 1, 3, 0, 0), 'weekday': 'Sunday', 'is_weekend': True, 'month': 1, 'Sweet_outlier': np.False_}
{'date': datetime.datetime(2021, 1, 4, 0, 0), 'weekday': 'Monday', 'is_weekend': False, 'month': 1, 'Sweet_outlier': np.False_}

Holiday flag counts: 17 true, 583 false
```

Conclusions, Observations, and Recommendations

Research provided significant and comprehensive findings about the data. The maximum times when people searched for diet-related products shown a clear sales increase in sweet products. The sales data establishes that people tend to follow dietary patterns that lead to hunger-related consumption. Practically useful information emerges from insights available to bakery business owners. Knowledge of dieting-related consumer behavior helps businesses optimize their inventory management while designing their marketing strategies and focusing promotional campaigns on changing seasonal dietary patterns.