

Министерство науки и высшего образования Российской Федерации  
Муромский институт (филиал)  
Федерального государственного бюджетного образовательного учреждения  
высшего образования  
«Владимирский государственный университет  
имени Александра Григорьевича и Николая Григорьевича Столетовых»

Факультет \_\_\_\_\_ ИТР \_\_\_\_\_

Кафедра \_\_\_\_\_ ПИН \_\_\_\_\_

## ЛАБОРАТОРНАЯ РАБОТА №4

По \_\_\_\_\_ Анализ данных \_\_\_\_\_  
Тема \_\_\_\_\_ Изучение функциональных возможностей  
\_\_\_\_\_ библиотеки Pandas \_\_\_\_\_  
\_\_\_\_\_

Руководитель

Белякова А.С.

\_\_\_\_\_  
(фамилия, инициалы)

\_\_\_\_\_  
(подпись)

\_\_\_\_\_  
(дата)

Студент \_\_\_\_\_ ПИН - 121 \_\_\_\_\_

\_\_\_\_\_  
(группа)

Ермилов М.В.

\_\_\_\_\_  
(фамилия, инициалы)

\_\_\_\_\_  
(подпись)

\_\_\_\_\_  
(дата)

Муром 2024

Цель работы: изучение возможностей Python для статистической обработки и визуализации данных с помощью библиотек Matplotlib, Pandas.

Задание: 1. Сгруппируйте транзакции по категориям, узнайте наиболее востребованные категории продуктов.

2. Выявите наиболее предпочтительный способ оплаты, зависит ли он от категории продуктов

3. Узнайте наиболее востребованные категории продуктов для различных категорий покупателей (сформируйте поло-возрастные группы 0-12 лет, 14-22 года, 23-35 лет, 26-50 лет, 50 – 65, 65 и старше). Варьируйте возраст и пол.

4. Какова востребованность использования кредитных карты при совершении покупок?

5. Выявите наиболее и наименее востребованные продукты, и долю затрат на них среди данной категории

6. Найдите самые дорогостоящие продукты и самые дешевые, 10 наиболее часто покупаемых и 10 наименее часто покупаемых

7. Выявите самые популярные торговые центры и узнайте, какие товары покупают в них чаще всего, реже всего.

8. Выявите лучшего покупателя (покупающего чаще всего, потратившего наибольшее количество денежных средств) каждого торгового центра.

Полученные результаты представьте графически. Выберите наиболее выразительные способ визуализации результатов статистической обработки. Оформите их убедительно и наглядно.

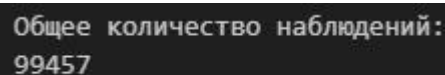
					МИВлГУ 09.03.04 - 0.09			
Изм.	Лист	№ докум.	Подпись	Дата				
Разраб.		Ермилов М.В.			Изучение функциональных возможностей библиотеки Pandas	Лит.	Лист	Листов
Провер.		Белякова А.С.					2	13
Реценз.						МИ ВлГУ ПИН-121		
Н. Контр.								
Утверд.								

Листинг кода:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#Указываем путь для чтения файла
path = ("C:/customer_shopping_data.csv") #Считываем файл
data = pd.read_csv(path)

#Отображаем информацию об общем количестве наблюдений
print("Общее количество наблюдений:")
print(data.shape[0])
```



```
Общее количество наблюдений:
99457
```

Рисунок 1 – результат работы программы

Задание 2.

# 1. Сгруппируйте транзакции по категориям, узнайте наиболее востребованные категории продуктов.

```
transactions_by_category =
data.groupby('category').size().sort_values(ascending=False).head(10)
```

```
# Визуализация наиболее востребованных категорий продуктов
plt.figure(figsize=(12, 6))
transactions_by_category.plot(kind='bar', color='skyblue')
plt.title('Top 10 Most Popular Product Categories')
plt.xlabel('Category')
plt.ylabel('Number of Transactions')
plt.show()
```

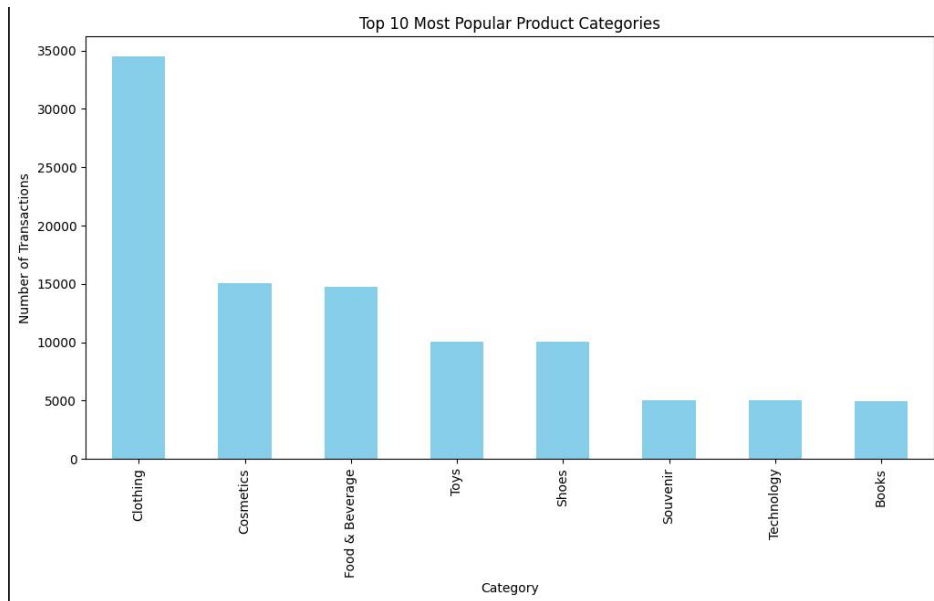


Рисунок 2 – результат работы программы

# 2. Выявите наиболее предпочтительный способ оплаты, зависит ли он от категории продуктов

```
preferred_payment_method = data.groupby('payment_method')['invoice_no'].count()
```

```
# Визуализация предпочтительного способа оплаты
plt.figure(figsize=(6, 6))
preferred_payment_method.plot(kind='pie', autopct='%1.1f%%',
colors=['lightgreen', 'lightblue', 'lightcoral'])
plt.title('Preferred Payment Method')
plt.ylabel("")
plt.show()
```

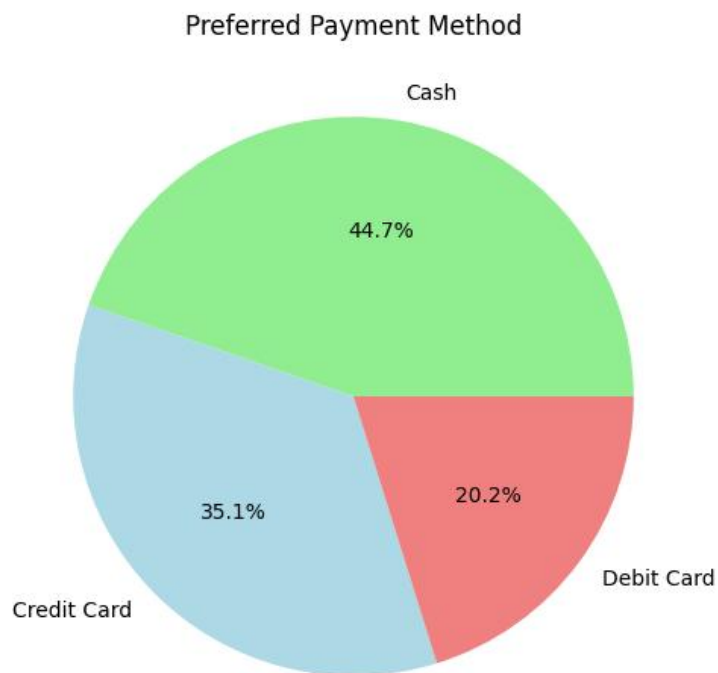


Рисунок 3 – результат работы программы

# 3. Узнайте наиболее востребованные категории продуктов для различных категорий покупателей

# Группировка данных по полу и возрастной категории

```
bins = [0, 12, 22, 35, 50, 65, 100]
```

```
labels = ['0-12', '14-22', '23-35', '26-50', '50-65', '65+']
```

```
data['age_group'] = pd.cut(data['age'], bins=bins, labels=labels, right=False)
```

# Построение графика

```
plt.figure(figsize=(12, 6))
```

```
for gender in data['gender'].unique():
```

```
    for age_group in data['age_group'].unique():
```

```
        grouped = data[(data['gender'] == gender) & (data['age_group'] == age_group)]
```

```
        category_count = grouped['category'].value_counts().nlargest(3) # Топ-3 наиболее востребованных категорий
```

```
plt.bar(x=gender + ' ' + age_group, height=category_count.values,
tick_label=category_count.index)
```

```
plt.xlabel('Gender and Age Group')
```

```
plt.ylabel('Count')
```

```
plt.title('Top Categories by Gender and Age Group')
```

```
plt.show()
```

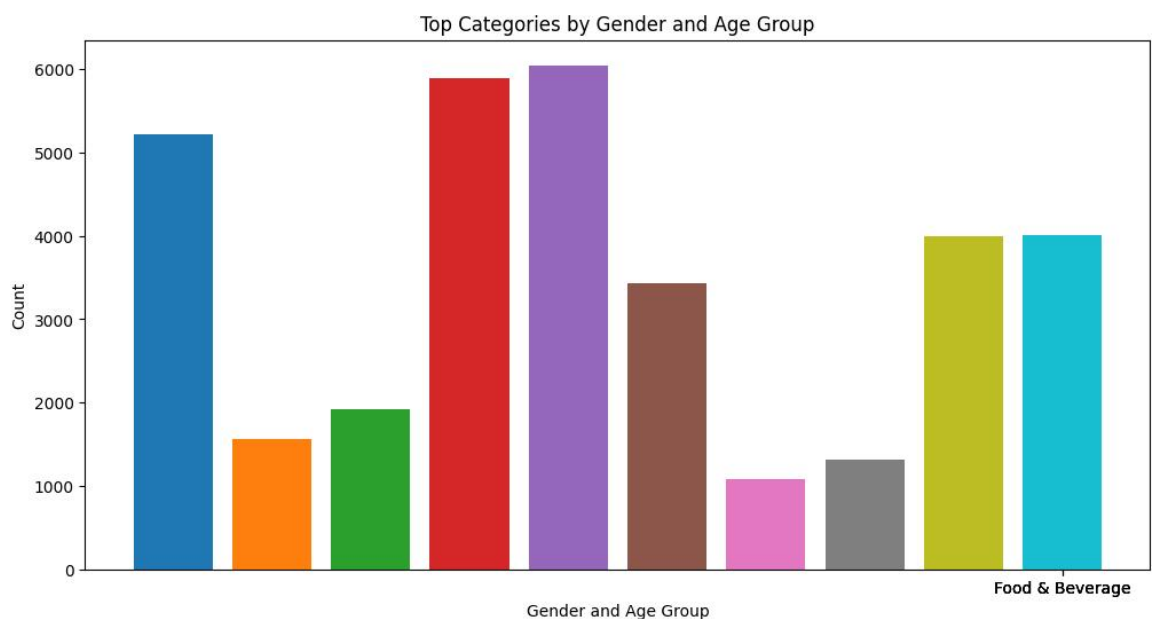


Рисунок 4 – результат работы программы

# 4. Какова востребованность использования кредитных карты при совершении покупок?

```
payment_method_category = data.groupby(['payment_method',
'category']).size().unstack()
```

# График предпочтительных способов оплаты по категориям

```
plt.figure(figsize=(12, 6))
```

```
payment_method_category.plot(kind='bar', stacked=True)
```

```
plt.title('Payment Methods by Product Category')
```

```
plt.xlabel('Product Categories')
```

```
plt.ylabel('Transaction Count')
plt.xticks(rotation=45)
plt.legend(title='Payment Method')
plt.show()
```

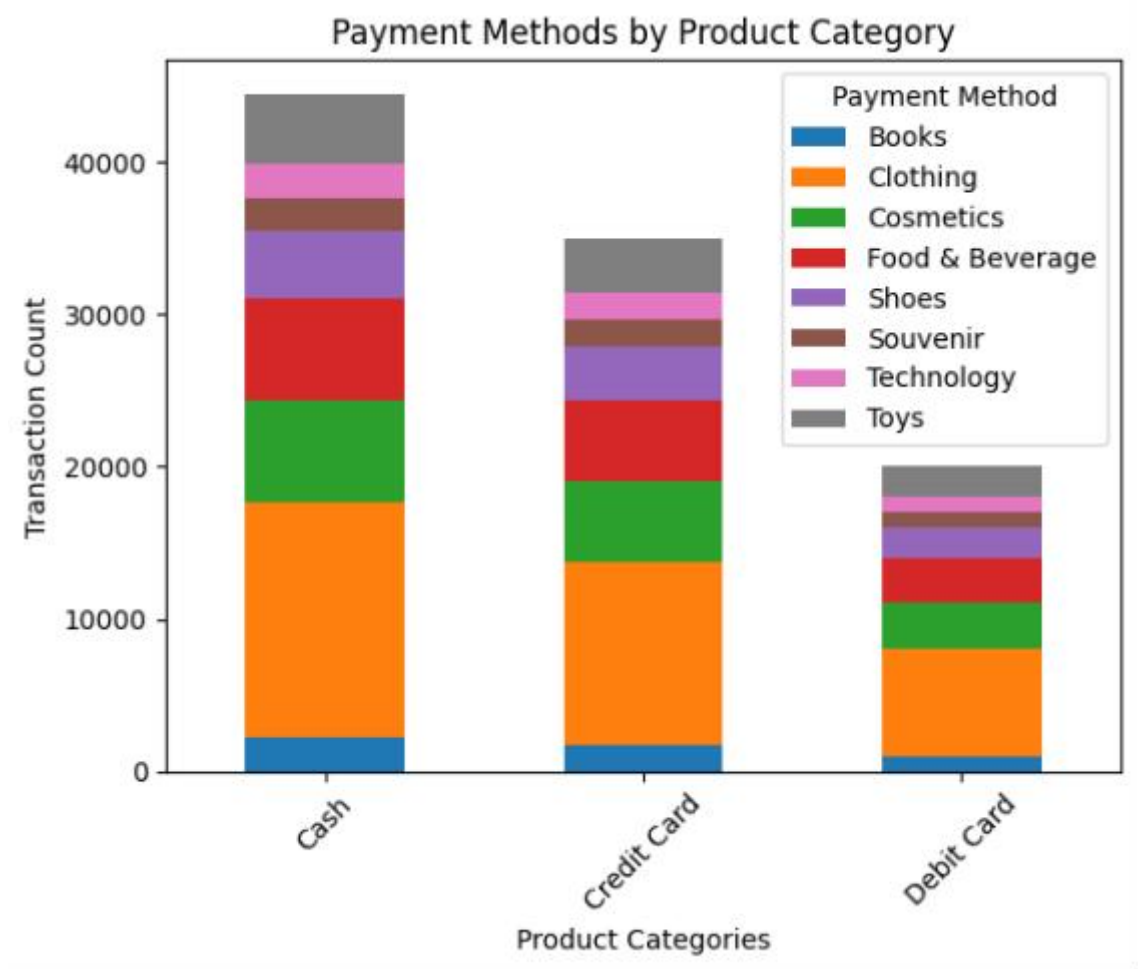


Рисунок 5 – результат работы программы

# 5. Выявите наиболее и наименее востребованные продукты, и их долю затрат в общей сумме

```
product_spending = data.groupby('category')['price'].sum()
most_demanding_products = product_spending.idxmax()
least_demanding_products = product_spending.idxmin()
product_share = product_spending / product_spending.sum()
# Визуализация доли затрат в общей сумме
plt.figure(figsize=(8, 6))
```

```
plt.bar(product_share.index, product_share.values, color='skyblue')
```

```
plt.xlabel('Category')
```

```
plt.ylabel('Share of Spending')
```

```
plt.title('Share of Spending by Category')
```

```
plt.xticks(rotation=45)
```

```
plt.show()
```

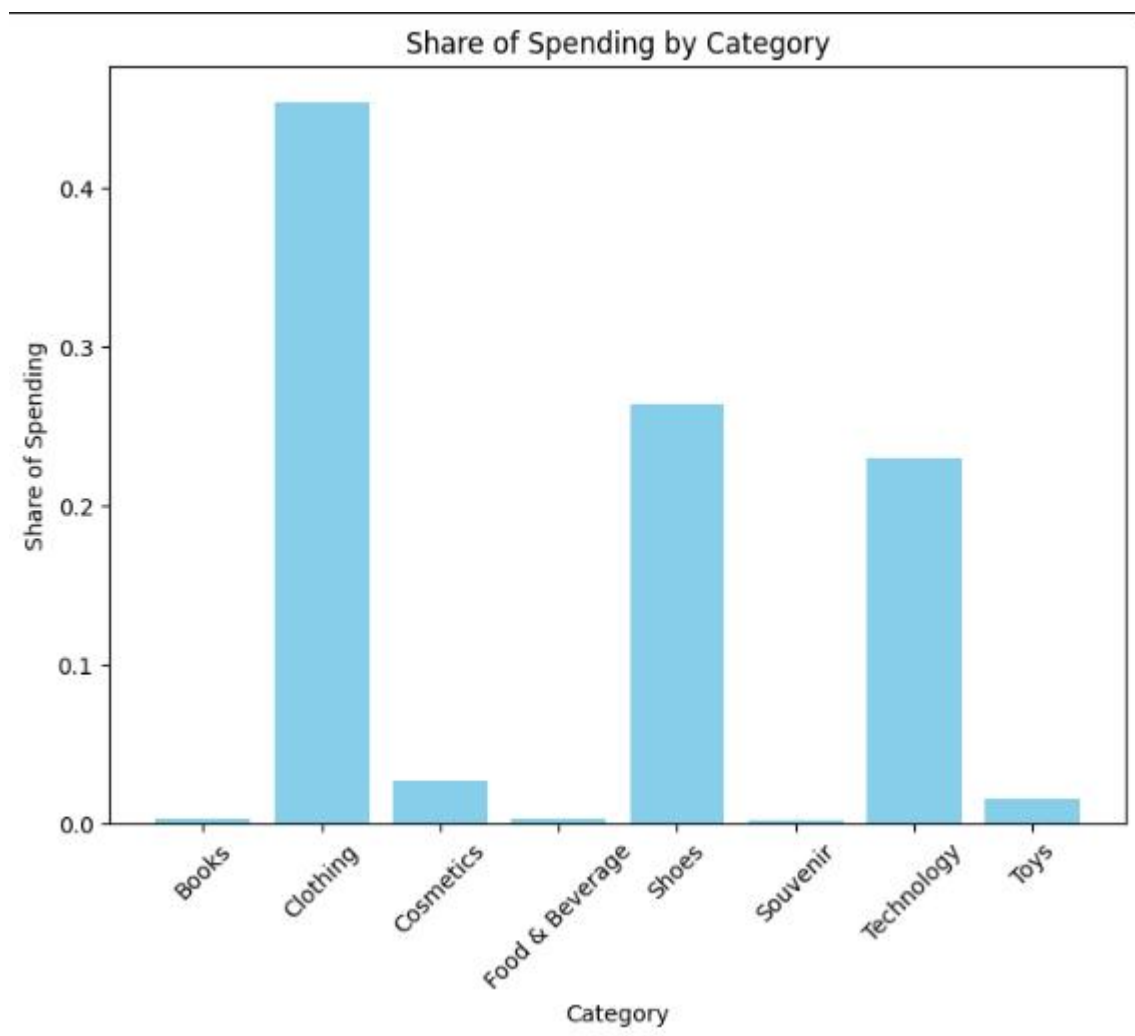


Рисунок 6 – результат работы программы

# 6. Найдите самые дорогостоящие продукты и самые дешевые, 10 наиболее часто покупаемых и 10 наименее часто покупаемых

# Нахождение самых дорогостоящих и самых дешевых продуктов



```

most_expensive_products = data.groupby('category')['price'].max().nlargest(1)
cheapest_products = data.groupby('category')['price'].min().nsmallest(1)

# Нахождение 10 наиболее и наименее популярных продуктов
most_popular_products = data['category'].value_counts().nlargest(10)
least_popular_products = data['category'].value_counts().nsmallest(10)

# Визуализация результатов
plt.figure(figsize=(12, 8))

plt.subplot(2, 2, 1)
plt.bar(most_expensive_products.index, most_expensive_products.values,
color='salmon')
plt.title('Most Expensive Product')

plt.subplot(2, 2, 2)
plt.bar(cheapest_products.index, cheapest_products.values, color='lightgreen')
plt.title('Cheapest Product')

plt.subplot(2, 2, 3)
plt.bar(most_popular_products.index, most_popular_products.values,
color='skyblue')
plt.title('Most Popular Products')
plt.xticks(rotation=45)

plt.subplot(2, 2, 4)
plt.bar(least_popular_products.index, least_popular_products.values,
color='lightcoral')
plt.title('Least Popular Products')
plt.xticks(rotation=45)

```

```
plt.tight_layout()
```

```
plt.show()
```

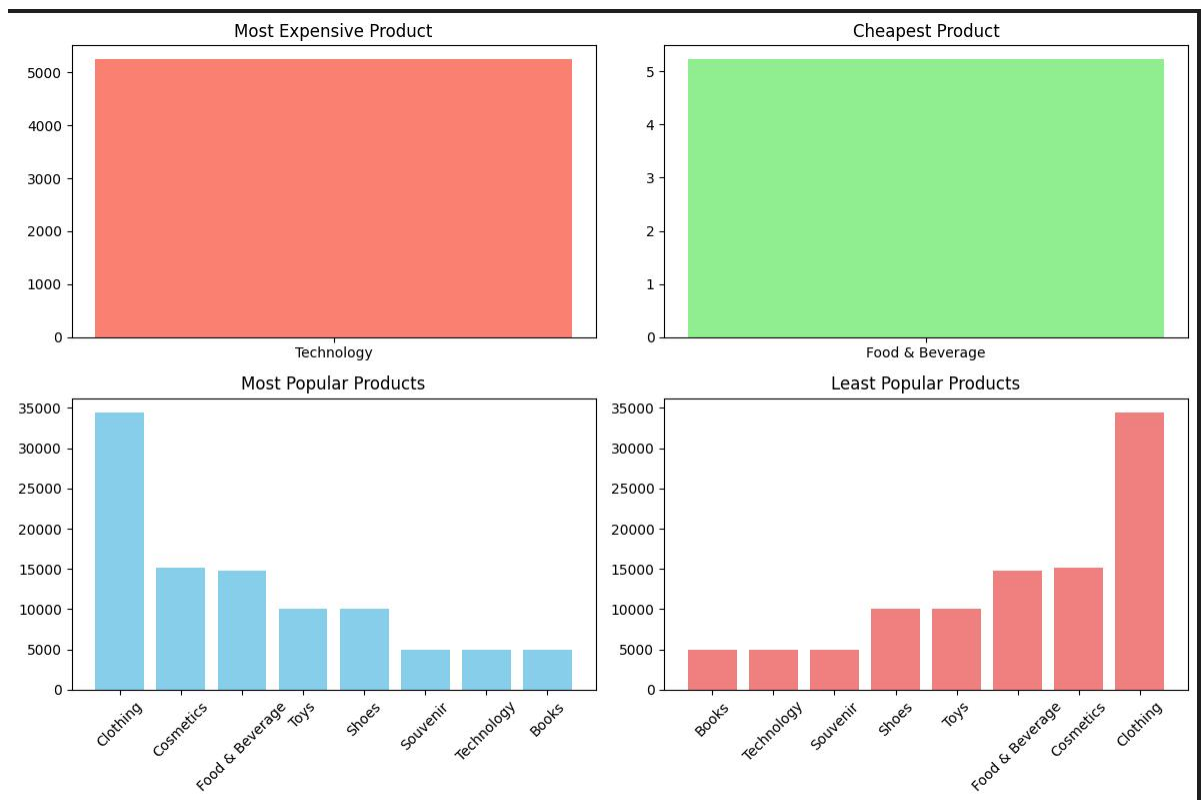


Рисунок 7 – результат работы программы

# 7. Выявите самые популярные торговые центры и узнайте, какие товары покупают в них чаще всего, реже всего

# Нахождение самых популярных и наименее популярных торговых центров

```
popular_malls = data['shopping_mall'].value_counts().nlargest()
```

```
rarely_visited_malls = data['shopping_mall'].value_counts().nsmallest()
```

# Нахождение самых популярных и наименее популярных категорий товаров для каждого торгового центра

```
popular_categories_per_mall =
```

```
data.groupby('shopping_mall')['category'].value_counts().groupby('shopping_mall').nlargest(1)
```

```

rare_categories_per_mall =
data.groupby('shopping_mall')['category'].value_counts().groupby('shopping_mall').nsmallest(1)

# Визуализация результатов
plt.figure(figsize=(12, 8))

plt.subplot(2, 2, 1)
plt.bar(popular_malls.index, popular_malls.values, color='lightblue')
plt.title('Popular Malls')

plt.subplot(2, 2, 2)
plt.bar(rarely_visited_malls.index, rarely_visited_malls.values,
color='lightcoral')
plt.title('Rarely Visited Malls')

plt.subplot(2, 2, 3)
plt.barh(popular_categories_per_mall.index.get_level_values('category'),
popular_categories_per_mall.values, color='lightgreen')
plt.title('Popular Categories per Mall')

plt.subplot(2, 2, 4)
plt.barh(rare_categories_per_mall.index.get_level_values('category'),
rare_categories_per_mall.values, color='lightsalmon')
plt.title('Rare Product Categories per Mall')

plt.tight_layout()
plt.show()

```

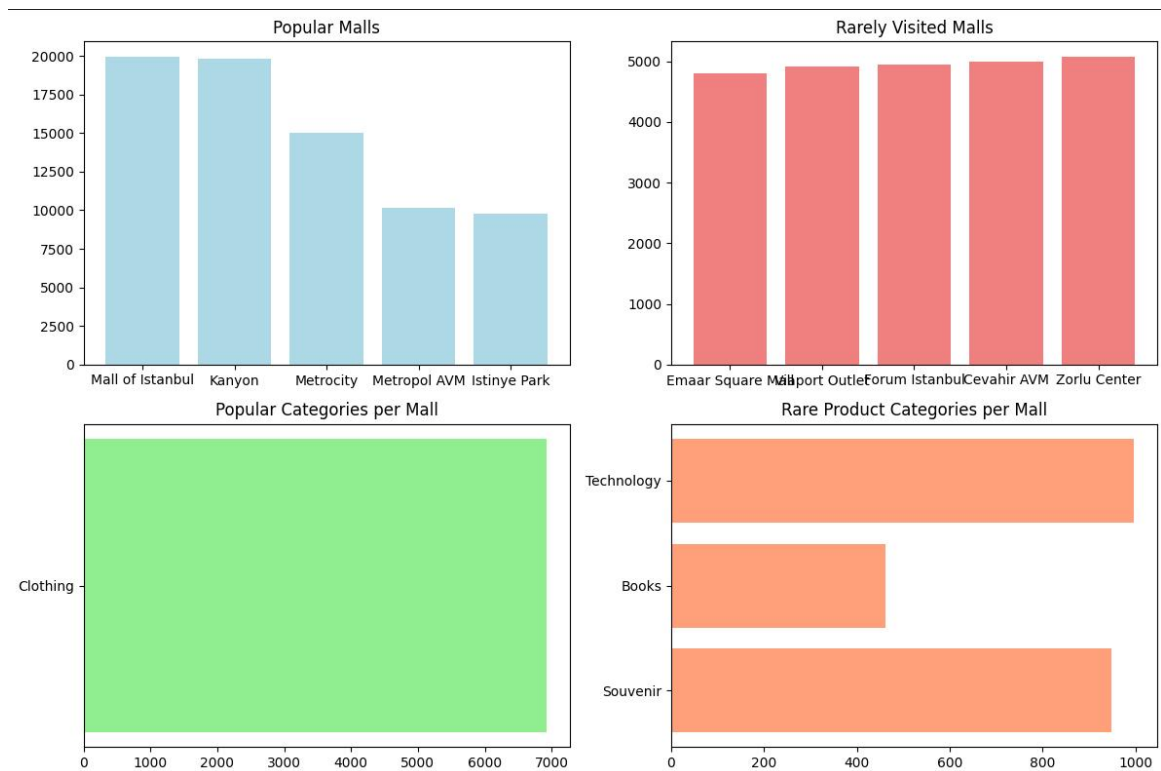


Рисунок 8 – результат работы программы

# 8. Выявите лучшего покупателя (покупающего чаще всего, потратившего наибольшее количество денежных средств) каждого торгового центра

# Группировка данных по торговым центрам и поиск лучшего покупателя

```
best_customer_per_mall = data.groupby('shopping_mall').apply(lambda x:
x.groupby('customer_id')['price'].sum().idxmax()).reset_index()
best_customer_per_mall.columns = ['shopping_mall', 'best_customer_id']
```

# Визуализация результатов

```
plt.figure(figsize=(12, 6))
plt.bar(best_customer_per_mall['shopping_mall'],
best_customer_per_mall['best_customer_id'], color='skyblue')
plt.xlabel('Shopping Mall')
plt.ylabel('Best Customer ID')
plt.title('Best Customer per Mall')
```

```
plt.xticks(rotation=45)
```

```
plt.show()
```

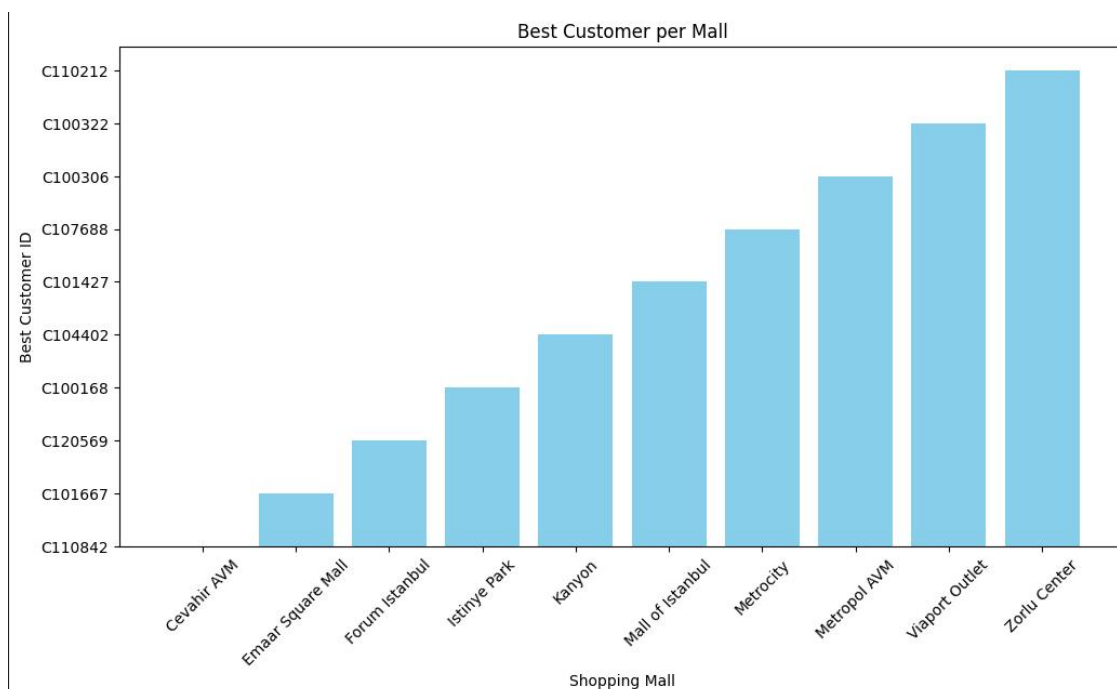


Рисунок 9 – результат работы программы

Вывод: в ходе лабораторной работы изучили возможности Python для статистической обработки и визуализации данных с помощью библиотек Matplotlib, Pandas.