

Міністерство освіти і науки України
Львівський національний університет імені Івана Франка

Факультет електроніки та комп'ютерних технологій

Звіт

Про виконання лабораторної роботи №2

3 курсу «Системи опрацювання даних»

ВІ аналітика

Business Intelligence Analyst

Виконав:

Студент групи ФеС-21

Осадчук Д.

Викладач:

Демків Л.С.

Львів-2025

Мета

З аналізу даних отримати відповіді на питання, які дозволять прийняти правильні рішення

Business intelligence (BI) — це збирання, зберігання і аналіз даних що утворюються при діяльності організації.¹ Метою business intelligence є підтримка прийняття кращих управлінських рішень

Завдання

1. Зчитати дані
2. Зрозуміти та написати про що дані. Відкинути непотрібні колонки.
3. Використовуючи відповідні методи pandas дослідити параметри даних. скільки і які це дані (числові чи категорії)

Придумати ще 10 питань до даних, які дозволять зрозуміти, що відбуваються з продажами для фінансового звіту. Для відповіді на питання написати відповідний код.

Наприклад: порівняти суму продажів всіх магазинів для кожного робочого дня тижня.

Порівняти продажі кожного одного магазину у робочі і вихідні дні.

Створити

Два з питань мають бути пов'язані з роботою з датами (наприклад, порівняти продажі за певний період)

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8
- 9
- 10

1. Зчитати і зрозуміти про що дані.

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

```
df = pd.read_csv('train.csv')
```

```
<ipython-input-4-cd75a77447fc>:1: DtypeWarning: Columns (7) have mixed types. Specify dtype
df = pd.read_csv('train.csv')
```

```
df.head()
```

	Store	DayOfWeek	Date	Sales	Customers	Open	Promo	StateHoliday	SchoolHoliday
0	1	5	2015-07-31	5263	555	1	1	0	1
1	2	5	2015-07-31	6064	625	1	1	0	1
2	3	5	2015-07-31	8314	821	1	1	0	1
3	4	5	2015-07-31	13995	1498	1	1	0	1
4	5	5	2015-07-31	4822	559	1	1	0	1

2. Використовуючи методи Pandas дослідити параметри даних

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1017209 entries, 0 to 1017208
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Store           1017209 non-null  int64
1   DayOfWeek       1017209 non-null  int64
2   Date            1017209 non-null  object
3   Sales           1017209 non-null  int64
4   Customers       1017209 non-null  int64
5   Open            1017209 non-null  int64
6   Promo           1017209 non-null  int64
7   StateHoliday    1017209 non-null  object
8   SchoolHoliday   1017209 non-null  int64
dtypes: int64(7), object(2)
```

df.describe()

	Store	DayOfWeek	Sales	Customers	Open	Promo	SchoolHoliday
count	1.017209e+06	1.017209e+06	1.017209e+06	1.017209e+06	1.017209e+06	1.017209e+06	1.017209e+06
mean	5.584297e+02	3.998341e+00	5.773819e+03	6.331459e+02	8.301067e-01	3.815145e-01	1.786467e-01
std	3.219087e+02	1.997391e+00	3.849926e+03	4.644117e+02	3.755392e-01	4.857586e-01	3.830564e-01
min	1.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
25%	2.800000e+02	2.000000e+00	3.727000e+03	4.050000e+02	1.000000e+00	0.000000e+00	0.000000e+00
50%	5.580000e+02	4.000000e+00	5.744000e+03	6.090000e+02	1.000000e+00	0.000000e+00	0.000000e+00
75%	8.380000e+02	6.000000e+00	7.856000e+03	8.370000e+02	1.000000e+00	1.000000e+00	0.000000e+00
max	1.115000e+03	7.000000e+00	4.155100e+04	7.388000e+03	1.000000e+00	1.000000e+00	1.000000e+00

	df.dtypes	
		0
	Store	int64
	DayOfWeek	int64
	Date	object
	Sales	int64
	Customers	int64
	Open	int64
	Promo	int64
	StateHoliday	object
	SchoolHoliday	int64

2. Придумати ще 10 питань до даних.

```
#1 Загальна сума від усіх продажів
total_sales = df['Sales'].sum()
print(f"Сума від усіх продажів: {total_sales}")
```

Сума від усіх продажів: 5873180623

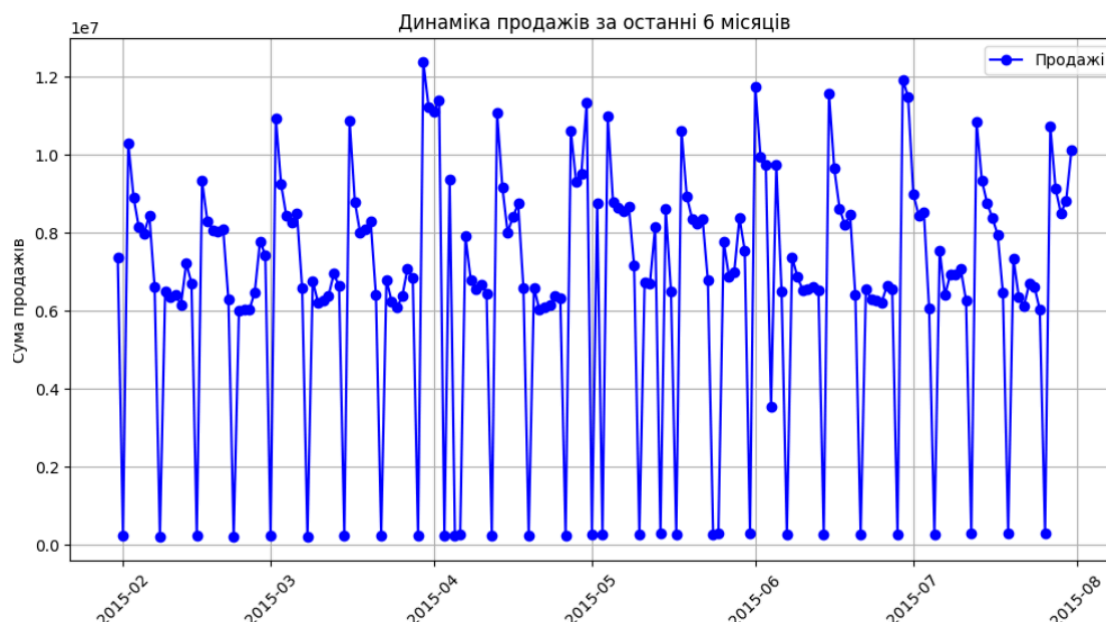
```
#2 Найменший продаж за добу
min_sales = df.groupby('Date')['Sales'].sum().min()
print(f"Найменший продаж за добу: {min_sales}")
```

Найменший продаж за добу: 97235

```
#3 Динаміка продажів за останні 6 міс.
df['Date'] = pd.to_datetime(df['Date'])

max_date = df['Date'].max()
six_months_ago = max_date - pd.DateOffset(months=6)
df_last_6_months = df[df['Date'] >= six_months_ago]
sales_trend = df_last_6_months.groupby('Date')['Sales'].sum()
plt.figure(figsize=(12, 6))
plt.plot(sales_trend.index, sales_trend.values, marker='o', linestyle='-', color='b', label='Продажі')

plt.title("Динаміка продажів за останні 6 місяців")
plt.xlabel("Дата")
plt.ylabel("Сума продажів")
plt.xticks(rotation=45)
plt.legend()
plt.grid()
plt.show()
```



```
#4 Середній продаж за добу
mean_sales = df.groupby('Date')['Sales'].sum().mean()
print(f"Середній продаж за добу: {mean_sales}")
```

Середній продаж за добу: 6234798.962845011

```
#5 ID магазину з найбільшими продажами
store_sales = df.groupby("Store")["Sales"].sum()
top_store = store_sales.idxmax()
print(f"Магазин з найбільшими продажами: {top_store}")
```

Магазин з найбільшими продажами: 262

```
#6 Список 5 магазинів з найб. продажами
top_sales_shop = df.groupby('Store')['Sales'].sum().sort_values(ascending=False)
print(top_sales_shop.head())
```

Store	
262	19516842
817	17057867
562	16927322
1114	16202585
251	14896870

Name: Sales, dtype: int64

```
#7 Середні продажі кожного магазину
avg_sales_per_store = df.groupby("Store")["Sales"].mean().reset_index()

print("\n Середні продажі кожного магазину:")
avg_sales_per_store.head(10)
```

Середні продажі кожного магазину:

	Store	Sales
0	1	3945.704883
1	2	4122.991507
2	3	5741.253715
3	4	8021.769639
4	5	3867.110403
5	6	4562.375796
6	7	7356.902335
7	8	4610.251592
8	9	5426.816348
9	10	4634.439490

```
#8 Список кількості продажів у вихідні та будні
df["Weekend"] = df["DayOfWeek"].apply(lambda x: 1 if x in [6, 7] else 0)
sales_weekday_vs_weekend = df.groupby("Weekend")["Sales"].mean().reset_index()
sales_weekday_vs_weekend["Weekend"] = sales_weekday_vs_weekend["Weekend"].map({0: "Будні", 1: "Вихідні"})

print(sales_weekday_vs_weekend)
```

	Weekend	Sales
0	Будні	6866.806351
1	Вихідні	3025.872894

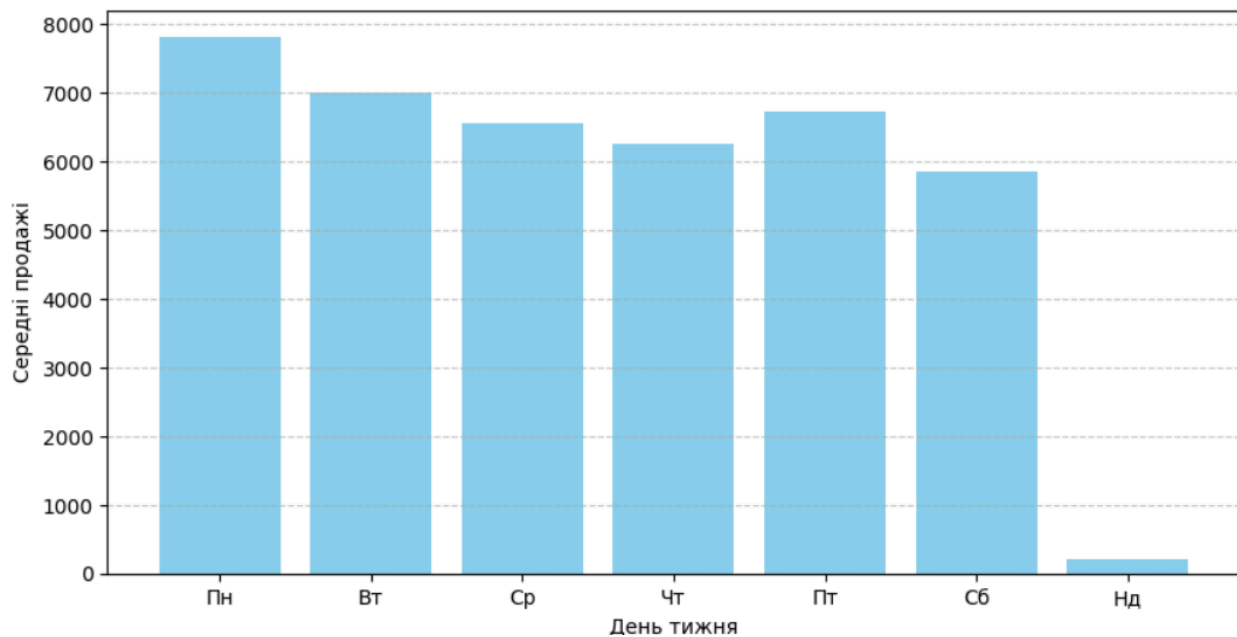
```
#10 Список дат коли кількість продажів була більшими за 12.000
monthly_sales = df.groupby('Date')['Sales'].mean().reset_index()
high_sales_months = monthly_sales[monthly_sales['Sales'] > 12000]

high_sales_months.head()
```

	Date	Sales
349	2013-12-16	14012.150673
356	2013-12-23	12870.005381
713	2014-12-15	12976.878075

```
#9 Середні продажі за кожен день тижня
avg_sales_per_day = df.groupby("DayOfWeek")["Sales"].mean().reset_index()

plt.figure(figsize=(10, 5))
plt.bar(avg_sales_per_day["DayOfWeek"], avg_sales_per_day["Sales"], color='skyblue')
plt.xlabel("День тижня")
plt.ylabel("Середні продажі")
plt.xticks(ticks=range(1, 8), labels=["Пн", "Вт", "Ср", "Чт", "Пт", "Сб", "Нд"])
plt.grid(axis="y", linestyle="--", alpha=0.7)
plt.show()
```



Висновок: під час другої лабораторної роботи я поглибив свої вміння у роботі з pandas, покращив свої навички у Business Intelligence.