

**Міністерство освіти і науки України Національний технічний університет України
«КПІ» імені Ігоря Сікорського Кафедра інформатики та програмної інженерії ФІОТ**

**ЗВІТ з лабораторної роботи №7 з навчальної дисципліни «Вступ до технології Data
Science»**

Тема:

**РОЗРОБКА ПРОГРАМНОГО МОДУЛЯ ПРОГНОЗУВАННЯ ДИНАМІКИ ЗМІНИ
ПОКАЗНИКІВ ЕФЕКТИВНОСТІ ТОРГІВЕЛЬНИХ КОМПАНІЙ
(міні проекти в галузі аналізу даних для завдань електронної комерції).**

Виконав

Студент 3 курсу кафедри ІІІ ФІОТ,
Навчальної групи ІІІ-12
Васильєв Є.К.

Перевірив

Професор кафедри ОТ ФІОТ
Писарчук О.О.

Київ 2023

I. Мета:

дослідити виявити та узагальнити особливості реалізації проектного практикуму в галузі аналізу часових (стохастичних рядів), як характеристика показників ефективності діяльності торговельних компаній.

II. Завдання:

4	Розробити програмний скрипт, що реалізує: 1. Парсинг файлу параметрів: Pr_1.xls; 2. Попередній аналіз даних; 3. Визначення показників ефективності – продаж та прибутку; 4. Визначення математичної моделі даних відповідно до МНК; 5. Здійснити прогнозування динаміки зміни продажів за регіонами (таблиця, графік).
---	---

III. Результати виконання лабораторної роботи.

3.1. Синтезована математична модель;

Метод найменших квадратів (МНК) - це статистичний метод, який використовується для апроксимації функцій шляхом мінімізації суми квадратів різниць між спостережуваними і очікуваними значеннями. Цей метод широко використовується в аналізі регресії та моделюванні для визначення параметрів функцій.

Основна ідея МНК полягає в тому, щоб знайти параметри моделі так, щоб сума квадратів відхилень (різниць) між спостережуваними значеннями і значеннями, розрахованими за допомогою моделі, була якомога меншою. Це досягається шляхом мінімізації суми квадратів різниць:

$$\sum_{i=1}^n (y_i - f(x_i, \beta))^2$$

Де n – кількість спостережень, (x_i, y_i) – координати кожного спостереження, $f(x_i, \beta)$ – значення функції з параметрами β у точці x_i , β – шукані параметри.

У випадку лінійної регресії, яка буде використовуватись для вирішення завдання, ця функція приймає вигляд:

$$f(x, \beta) = \beta_i * x + \beta_0$$

Де β_i – нахил (slope), β_0 – перетин (intercept).

3.2. Результати архітектурного проектування та їх опис;

Після дослідження структури файлу та його парсингу, було виділено два датафрейми: з даними про продажі, та списком усіх регіонів. Отримані датафрейми було розглянуто, та в результаті попереднього аналізу було змінено назву некоректного місяця, перейменовано стовпці й перевірено співпадіння кодів регіонів. Також до датафрейму з продажами було додано стовпець виторгу (кількість реалізацій * ціна реалізації) та прибутку ((ціна реалізації – собівартість одиниці) * кількість реалізацій). Після цього було візуалізовано виторг та прибуток по місяцям, визначено математичну модель відповідно до МНК та здійснено прогнозування динаміки зміни продажів за регіонами. Результати було відтворено за допомогою графіку та таблиці.

3.3. Опис структури проекту програми;



Рис.1. Блок схема алгоритму програми.

3.4. Результати роботи програми відповідно до завдання (допускається у формі скриншотів);

```
['Січень' 'Лютий' 'Березень' 'Квітень' 'Травень' 'Червень' 'Липень'  
'Серпень' 'Вересень' 'Жовтень' 'Грудень' 'Листопад' 'Стоп']
```

Рис.2. Перевірка коректності назв місяців.

```
Місяць  
Січень      282941.242021  
Лютий       317555.150758  
Березень    223312.168412  
Квітень     179349.838641  
Травень     189906.073095  
Червень     204735.325974  
Липень      276280.556562  
Серпень     309607.183185  
Вересень    276352.358834  
Жовтень     278810.546447  
Листопад    296732.726115  
Грудень     433293.527863  
Name: Виторг, dtype: float64  
Місяць  
Січень      6302.837821  
Лютий       -35441.849242  
Березень    63411.168412  
Квітень     33292.838641  
Травень     40182.073095  
Червень     34404.325974  
Липень      36488.556562  
Серпень     49418.183185  
Вересень    32308.366134  
Жовтень     38886.393947  
Листопад    26657.517515  
Грудень     47081.816563  
Name: Прибуток, dtype: float64
```

Рис.3. Обрахунок загальних продажів та прибутку.

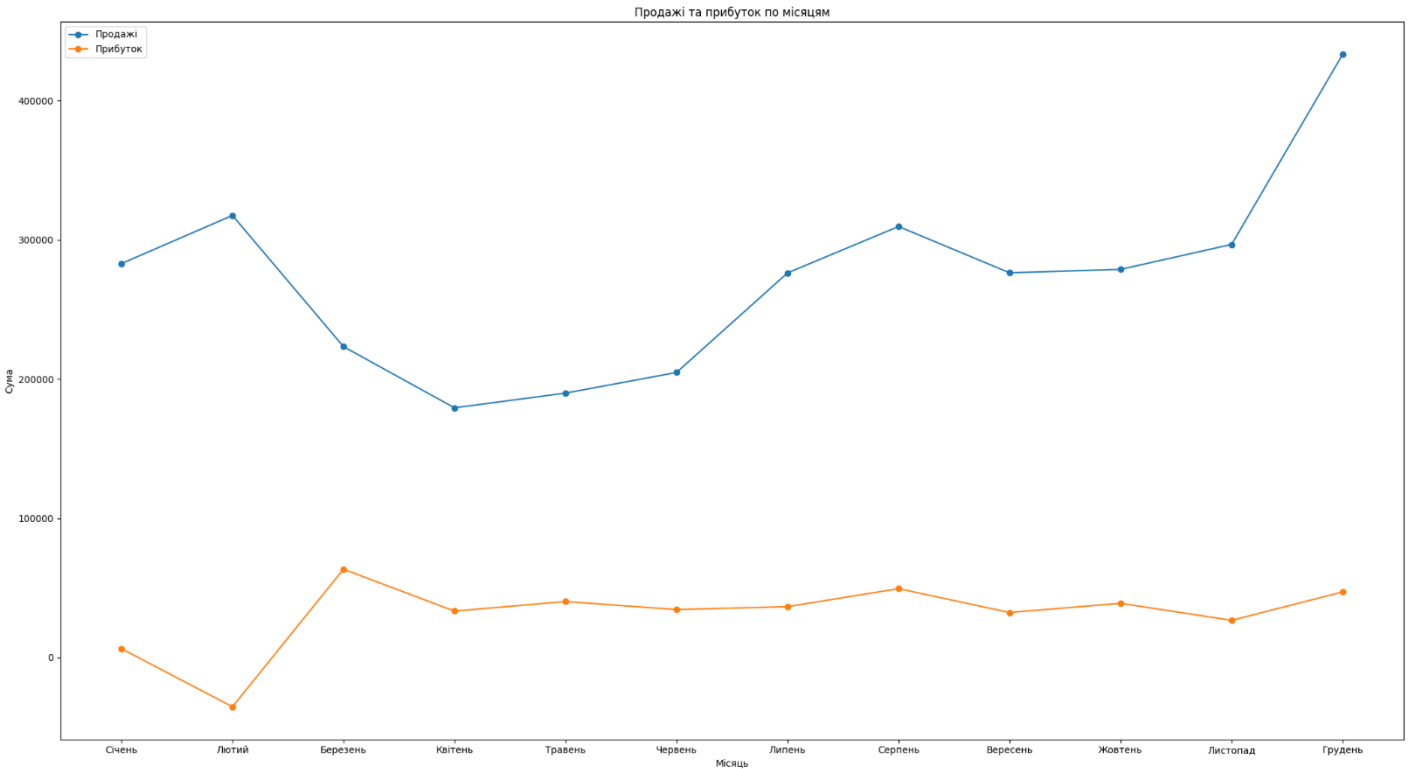


Рис.4. Візуалізація продажів та прибутку.

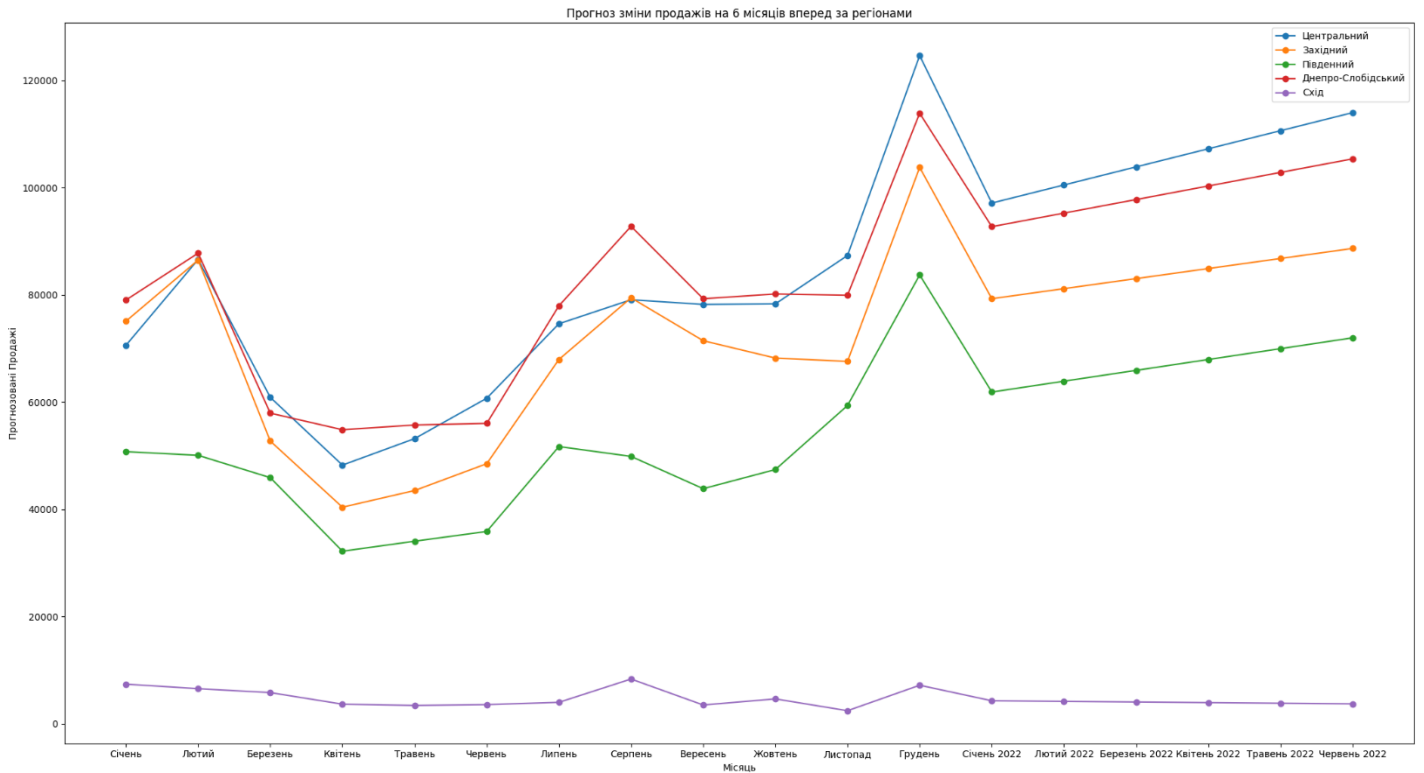


Рис.5. Прогнозування зміни продажів на пів року вперед.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Січень	Лютий	Березень	Квітень	Травень	Червень	Липень	Серпень	Вересень	Жовтень	Листопад	Грудень	Січень 2022	Лютий 2022	Березень 2022	Квітень 2022	Травень 2022	Червень 2022
2	79059,70814	87798,50996	57951,69282	54853,64473	55739,86405	56040,6381	77976,79228	92804,41234	79296,91806	80189,07198	79948,0649	113907,2636	92744,4179	95274,7568	97805,09569	100335,4346	102865,7735	105396,1124
3	75114,69499	86513,78285	52739,30959	40399,30633	43511,04524	48518,25238	67965,12554	79470,38788	71465,04622	68226,68139	67607,24254	103800,7184	79303,0706	81178,77898	83054,48737	84930,19575	86805,90413	88681,61252
4	50765,41938	50079,95121	45915,35348	32190,55805	34051,8031	35876,87879	51714,1316	49881,30519	43858,4881	47423,96655	59377,4257	83736,39886	61888,93725	63911,95734	65934,97743	67957,99752	69981,01761	72004,0377
5	7385,731964	6538,736568	5810,322831	3648,723217	3410,225714	3568,124026	3993,891991	8341,175758	3496,786007	4632,806851	2429,650839	7183,023762	4286,708106	4171,340128	4055,972151	3940,604173	3825,236196	3709,868218
6	70615,68754	86624,17016	60895,4897	48257,60631	53193,135	60731,43268	74630,61515	79109,90202	78235,12044	78338,01967	87370,34214	124666,1233	97151,96515	100525,7592	103899,5533	107273,3474	110647,1414	114020,9355

Рис.6. Збереження вихідних даних в Excel таблицю.

3.5. Програмний код, що забезпечує отримання результату (допускається у формі скриншотів).

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 pd.set_option('display.max_columns', None)
6
7 sales_df = pd.read_excel(r'C:\Users\egorv\PycharmProjects\pythonProject\Data science\Lab7\Example\Pr_1.xls', usecols=[x for x in range(6)])
8 store_locations_df = pd.read_excel(r'C:\Users\egorv\PycharmProjects\pythonProject\Data science\Lab7\Example\Pr_1.xls', usecols=[9, 10], nrows=199)
9 sales_df.rename(columns={'Місяць': 'Місяць', 'Кількість реалізацій': 'Кількість реалізацій', 'Ціна реалізації': 'Ціна реалізації'}, inplace=True)
10 store_locations_df.rename(columns={'Код магазину.1': 'Код магазину'}, inplace=True)
11
12 print(sales_df['Місяць'].unique())
13 print(sales_df[sales_df['Місяць'] == 'Стон'])
14 sales_df['Місяць'] = sales_df['Місяць'].replace('Стон', 'Листопад')
15 print(store_locations_df['Код магазину'].isin(sales_df['Код магазину'].unique()).all())
16
17 sales_df['Виторг'] = sales_df['Кількість реалізацій'] * sales_df['Ціна реалізації']
18 sales_df['Прибуток'] = (sales_df['Ціна реалізації'] - sales_df['Собівартість одиниці']) * sales_df['Кількість реалізацій']
19 monthly_sales = sales_df.groupby('Місяць')['Виторг'].sum()
20 monthly_profit = sales_df.groupby('Місяць')['Прибуток'].sum()
21 months_order = ['Січень', 'Лютий', 'Березень', 'Квітень', 'Травень', 'Червень', 'Липень', 'Серпень', 'Вересень', 'Жовтень', 'Листопад', 'Грудень']
22 monthly_sales = monthly_sales.reindex(months_order)
23 monthly_profit = monthly_profit.reindex(months_order)
24 print(monthly_sales, monthly_profit)
25 plt.figure(figsize=(10, 6)) # plotting sales and profit by months
26 plt.plot(monthly_sales.index, monthly_sales, label='Продажі', marker='o')
27 plt.plot(monthly_profit.index, monthly_profit, label='Прибуток', marker='o')
28 plt.xlabel('Місяць')
29 plt.ylabel('Сума')
30 plt.title('Продажі та прибуток по місяцям')
31 plt.legend()
32 plt.show()
33
34 df_merged = sales_df.merge(store_locations_df, on='Код магазину', how='left')
35 regions = df_merged['Perion'].unique()
36 df_merged['Місяць цифра'] = df_merged['Дата'].dt.month
37 months = np.sort(df_merged['Місяць цифра'].unique())
38 forecast_df = pd.DataFrame(index=regions, columns=[f'{month} 2022' for month in months_order[:6]]) # DataFrame for storing forecasts
39 df_pivot = df_merged.pivot_table(values='Виторг', index='Perion', columns='Місяць', aggfunc="sum") # Pivoting the DataFrame for ease of use
40 df_pivot = df_pivot[months_order]
41 # Defining forecasts for the next 6 months for each region
42 for region in regions:
43     region_sales = df_pivot.loc[region, months_order]
44     coefficients = np.polyfit(months, region_sales, 1)
45     forecast_sales = np.polyval(coefficients, months[-1] + np.arange(1, 7))
46     forecast_df.loc[region] = forecast_sales
47 sales_and_forecast_df = pd.concat([df_pivot, forecast_df], axis=1) # Concatting dataframe with actual sales and predicted
48 plt.figure(figsize=(12, 8))
49 for region in regions:
50     p = sales_and_forecast_df.loc[region]
51     plt.plot(sales_and_forecast_df.loc[region], label=region, marker='o')
52 plt.xlabel('Місяць')
53 plt.ylabel('Прогнозовані Продажі')
54 plt.title('Прогноз зміни продажів на 6 місяців вперед за регіонами')
55 plt.legend()
56 plt.show()
57
58 sales_and_forecast_df.to_excel('Output.xlsx', index=False)
```

IV. Висновки.

В ході лабораторної роботи було розроблено програмний скрипт, що реалізує парсинг, аналіз та прогнозування даних з вхідного Excel файлу.

Досліджуючи динаміку загальних продажів та прибутку, було встановлено, що хоч виторг і мав позитивний тренд починаючи з квітня 21го року, заробіток при цьому залишався стабільним, з чого можна зробити припущення, що інтенсивна конкуренція на ринку може тиснути на компанію, змушуючи її підтримувати конкурентоспроможні ціни або пропонувати рекламні акції, щоб зберегти або збільшити свою долю на ринку. Однак щоб зробити більш точний висновок необхідно детальніше проаналізувати фінансову звітність, дискусії керівництва та ринкові умови, аби зрозуміти конкретну динаміку компанії.

Для прогнозування обсягів продажів було обрано саме лінійну регресію, оскільки вхідні дані не мають складної структури та демонструють пряму залежність. Крім того вона дає змогу чітко виявити тренд та зрозуміти загальну динаміку.

Прогнозування продажів, яке вказує на позитивний тренд, дозволяє підготувати запаси для зростаючого попиту. Це важливо для ефективних закупівель та уникнення втрат прибутку, однак реальний бізнес має багато нюансів, які навіть найточніші моделі не завжди здатні врахувати. Економічні, конкурентні та інші зміни можуть суттєво впливати на стратегію. Тому важливо об'єднувати прогнози з гнучкістю та адаптацією до реальних умов. Тільки цей підхід гарантує стійкість та успішність бізнесу в довгостроковій перспективі.

Отримані результати та побудована модель мають безпосереднє практичне значення для бізнесу. Розуміння та використання подібних методів аналізу та прогнозування дозволяє приймати обґрунтовані управлінські рішення, підвищуючи ефективність бізнес-процесів та відповідаючи на змінні умови ринку.

Виконав: Васильєв Єгор