**Міністерство освіти і науки України**

**Національний технічний університет України «КПІ» імені Ігоря Сікорського**

**Кафедра обчислювальної техніки ФІОТ**

**ЗВІТ**

**з лабораторної роботи №7**

**з навчальної дисципліни «Вступ до технології Data Science»**

**Тема:**

**ДОСЛІДЖЕННЯ МЕТОДІВ ПРОГНОЗУВАННЯ ДИНАМІКИ ЗМІН ПОКАЗНИКІВ ЕФЕКТИВНОСТІ ТРЕЙДИНГОВИХ КОМПАНІЙ**

**Виконав:**

Студент X курсу кафедри ОТ ФІОТ,

Навчальної групи ІА-73

Петренко П.П.

**Перевірив:**

Професор кафедри ОТ ФІОТ

Писарчук О.О.

**Київ 2023**

**І. Мета:**

виявити дослідити та узагальнити особливості застосування методів прогнозування динаміки змін показників ефективності трейдингових компаній з використанням спеціалізованих пакетів мови програмування Python.

**ІІ. Завдання:**

**Завдання.**

Лабораторія провідної ІТ-компанії реалізує проект із розробки системи підтримки прийняття рішень для автоматизації процесів вироблення рішень у сфері задач електронної комерції.

Замовниками програмної системи – низка оптових компаній.

Вам, як Data Science [Engineer](https://jobs.dou.ua/companies/gravitum/vacancies/147764/) поставлене наступне завдання.

Розробити скрипт мовою Python що реалізує прогнозування динаміки зміни показників ефективності трейдингових компаній відповідно до технічних умов, заданих у таблиці додатку 1.

Завдання реалізувати у відповідності до пунктів:

* 1. Обрати варіант відповідно до рівня складності;
  2. Здійснити аналіз даних відповідно до технічних умов обраного варіанту.
  3. Здійснити візуалізацію результатів аналізу (у формі таблиці та графіків);
  4. Оцінити ефективність розробленого скрипта. Оцінювання ефективності здійснити за методичними та статистичними характеристиками отриманих результатів.

Результат представити у формі:

1.5. Результати архітектурного проектування скрипта, що реалізує технічни умови задачі.

1.6. Програмний скрипт, результати його функцілнування.

1.7. Результати візуалізації процесу прогнозування динаміки зміни показників ефективності трейдингових компаній.

| 7, 22 | Розробити програмний скрипт, що реалізує:  1. Парсінг файлу параметрів: Pr12.csv  2. Попередній аналіз даних;  3. Визначення показників ефективності – продажи та прибутку;  4. Визначення математичної моделі даних відповідно до МНК;  5. Визначити динаміку зміни продажів за регіонами (таблиця, графік, файл). |
| --- | --- |

| 7, 22 | Розробити програмний скрипт, що реалізує аналіз даних, поданих у файлі Data\_Set\_7.csv |
| --- | --- |

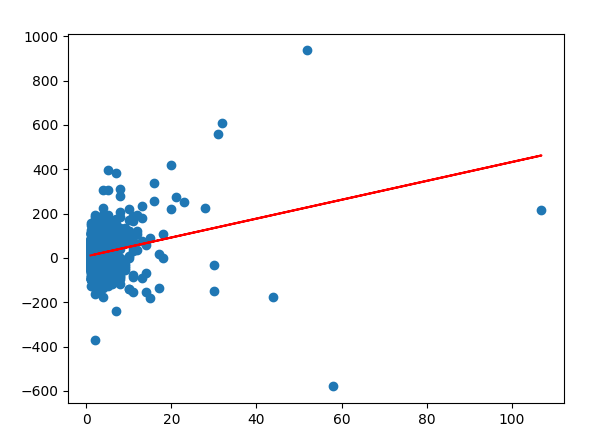


Рисунок 1. Прогнозування динаміки зміни прибутку

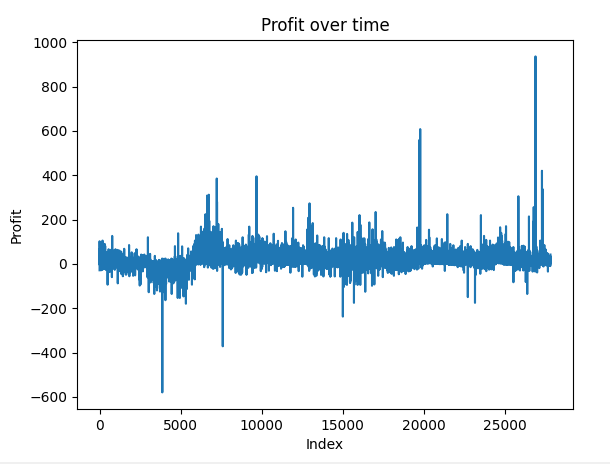


Рисунок 2 Зміна прибутку

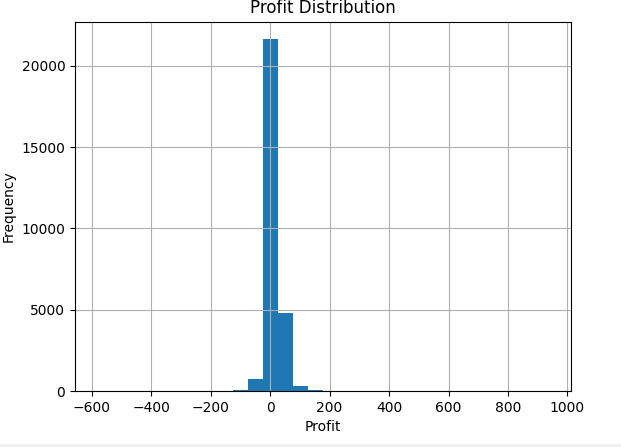


Рисунок 3. Розподіл значень прибутку

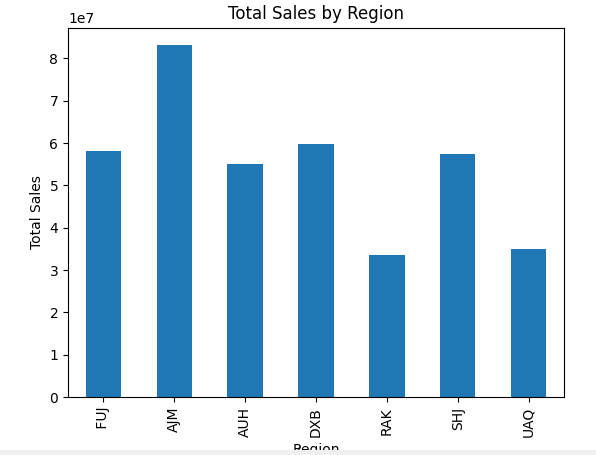


Рисунок 4. Розподіл продажів по регіонам

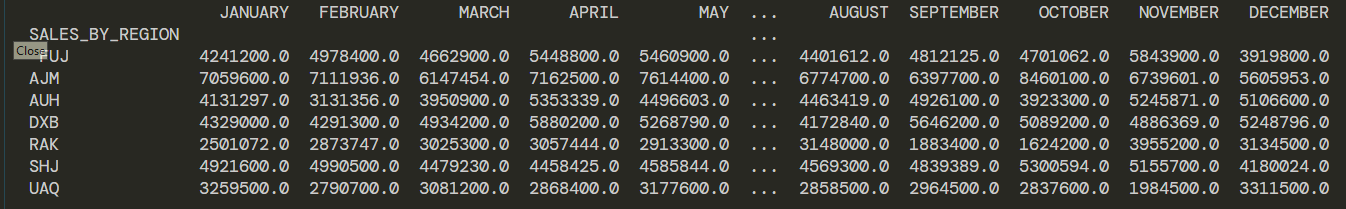


Рисунок 5. Розподіл продажів по регіонам та місяцям

**ІІІ. Довести адекватність сформованих моделей та працездатність розробленого скріпта.**

**Розроблений код повинен бути раціональним та відповідати вимогам до чистого коду.**

***import* pandas *as* pd**

***import* numpy *as* np**

***import* matplotlib.pyplot *as* plt**

***from* sklearn.linear\_model *import* LinearRegression**

**df *=* pd.read\_csv("Pr12\_2.csv")**

**df.info()**

**df.describe()**

**df.head()**

**df['Number of sales'] *=* pd.to\_numeric(df['Number of sales'], errors*=*'coerce')**

**df *=* df.dropna(subset*=*['Number of sales'])**

***# Calculate profit***

**df['Profit'] *=* df['Number of sales'] *\** (df['Selling price'] *-* df['Unit cost'])**

**df *=* df.dropna()**

**df['Number of sales'] *=* pd.to\_numeric(df['Number of sales'], errors*=*'coerce')**

**df['Profit'] *=* df['Number of sales'] *\** (df['Selling price'] *-* df['Unit cost'])**

**model *=* LinearRegression()**

**model.fit(df['Number of sales'].values.reshape(*-*1,1), df['Profit'])**

**predictions *=* model.predict(df['Number of sales'].values.reshape(*-*1,1))**

**plt.scatter(df['Number of sales'], df['Profit'])**

**plt.plot(df['Number of sales'], predictions, color*=*'red')**

**plt.show()**

**df['Number of sales'] *=* pd.to\_numeric(df['Number of sales'], errors*=*'coerce')**

**df['Profit'] *=* df['Number of sales'] *\** (df['Selling price'] *-* df['Unit cost'])**

**df['Profit'].plot(kind*=*'line')**

**plt.title('Profit over time')**

**plt.xlabel('Index')**

**plt.ylabel('Profit')**

**plt.show()**

**df['Profit'].hist(bins*=*30)**

**plt.title('Profit Distribution')**

**plt.xlabel('Profit')**

**plt.ylabel('Frequency')**

**plt.show()**

***# 7, 22 Розробити програмний скрипт, що реалізує аналіз даних, поданих у файлі Data\_Set\_7.csv***

***import* matplotlib.pyplot *as* plt**

***import* pandas *as* pd**

***import* numpy *as* np**

***# Parse the CSV file***

**df *=* pd.read\_csv("Data\_Set\_6.csv")**

**df *=* df.replace(['n.a.', 'not avilable'], np.nan)**

***for* month *in* ['JANUARY', 'FEBRUARY', 'MARCH', 'APRIL', 'MAY', 'JUNE', 'JULY', 'AUGUST', 'SEPTEMBER', 'OCTOBER', 'NOVEMBER', 'DECEMBER']:**

**df[month] *=* df[month].astype(str).str.replace(*r*'\D', '', regex*=*True).replace('', np.nan).astype(float)**

**total\_sales\_by\_region *=* df.groupby('SALES\_BY\_REGION')[['JANUARY', 'FEBRUARY', 'MARCH', 'APRIL', 'MAY', 'JUNE', 'JULY', 'AUGUST', 'SEPTEMBER', 'OCTOBER', 'NOVEMBER', 'DECEMBER']].sum()**

**print(total\_sales\_by\_region)**

**total\_sales\_by\_region *=* df.groupby('SALES\_BY\_REGION')[['JANUARY', 'FEBRUARY', 'MARCH', 'APRIL', 'MAY', 'JUNE', 'JULY', 'AUGUST', 'SEPTEMBER', 'OCTOBER', 'NOVEMBER', 'DECEMBER']].sum()**

**total\_sales\_by\_region['TOTAL\_SALES'] *=* total\_sales\_by\_region.sum(axis*=*1)**

**total\_sales\_by\_region['TOTAL\_SALES'].plot(kind*=*'bar')**

**plt.title('Total Sales by Region')**

**plt.xlabel('Region')**

**plt.ylabel('Total Sales')**

**plt.show()**

**III. Висновок:**

виявив дослідив та узагальнив особливості застосування методів прогнозування динаміки змін показників ефективності трейдингових компаній з використанням спеціалізованих пакетів мови програмування Python.

Виконав: студент Петренко П.П.