

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ

імені ІГОРЯ СІКОРСЬКОГО”

Факультет прикладної математики

Кафедра програмного забезпечення комп’ютерних систем



**Лабораторна робота №4**

з дисципліни: «Технології оброблення великих даних»

на тему: «Побудова лінійної регресії в Python»

Виконав

студент IІІ курсу каф. ПЗКС ФПМ

групи КП-82

Мельничук Олексій Геннадійович

Перевірила

доц. каф. ПЗКС ФПМ

Олещенко Л.М.

Київ 2021

1. **Індивідуальне завдання**

**Мета**: ознайомитись з поняттями простої лінійної регресії та роботи з наданими даними для прогнозування в Python.

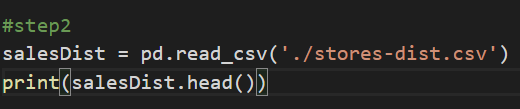
У статистиці лінійна регресія - це спосіб моделювання взаємозв'язку між залежною змінною р і незалежною змінною х. У цій лабораторній роботі ви проаналізуєте дані про продажі району та побудуєте просту лінійну регресію для прогнозування річного чистого обсягу продажів на основі кількості магазинів у районі.

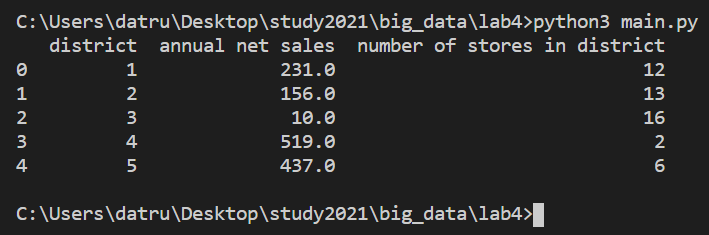
1. **Хід роботи**

|  |
| --- |
| **Stores\_dist.txt** |
| Gender,FSIQ,VIQ,PIQ,Weight,Height,MRI\_Count  Female,133,132,124,118,64.5,816932  Male,140,150,124,NaN,72.5,1001121  Male,139,123,150,143,73.3,1038437  Male,133,129,128,172,68.8,965353  Female,137,132,134,147,65.0,951545  Female,99,90,110,146,69.0,928799  Female,138,136,131,138,64.5,991305  Female,92,90,98,175,66.0,854258  Male,89,93,84,134,66.3,904858  Male,133,114,147,172,68.8,955466  Female,132,129,124,118,64.5,833868  Male,141,150,128,151,70.0,1079549  Male,135,129,124,155,69.0,924059  Female,140,120,147,155,70.5,856472  Female,96,100,90,146,66.0,878897  Female,83,71,96,135,68.0,865363  Female,132,132,120,127,68.5,852244  Male,100,96,102,178,73.5,945088  Female,101,112,84,136,66.3,808020  Male,80,77,86,180,70.0,889083  Male,83,83,86,NaN,NaN,892420  Male,97,107,84,186,76.5,905940  Female,135,129,134,122,62.0,790619  Male,139,145,128,132,68.0,955003  Female,91,86,102,114,63.0,831772  Male,141,145,131,171,72.0,935494  Female,85,90,84,140,68.0,798612  Male,103,96,110,187,77.0,1062462  Female,77,83,72,106,63.0,793549  Female,130,126,124,159,66.5,866662  Female,133,126,132,127,62.5,857782  Male,144,145,137,191,67.0,949589  Male,103,96,110,192,75.5,997925  Male,90,96,86,181,69.0,879987  Female,83,90,81,143,66.5,834344  Female,133,129,128,153,66.5,948066  Male,140,150,124,144,70.5,949395  Female,88,86,94,139,64.5,893983  Male,81,90,74,148,74.0,930016  Male,89,91,89,179,75.5,935863 |

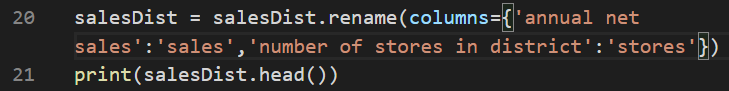
**Частина 1: Імпорт бібліотек та даних**

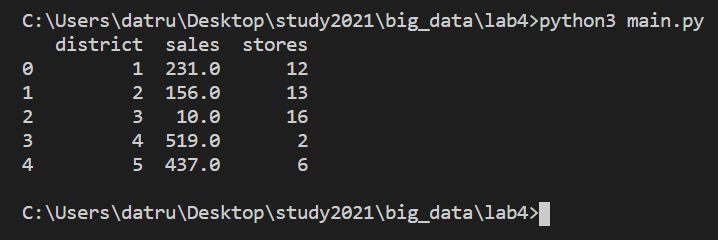
**Крок 2: Імпортуйте дані.**





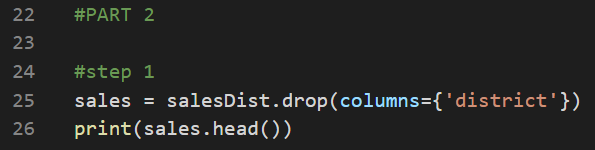
Заголовки імпортовано правильно

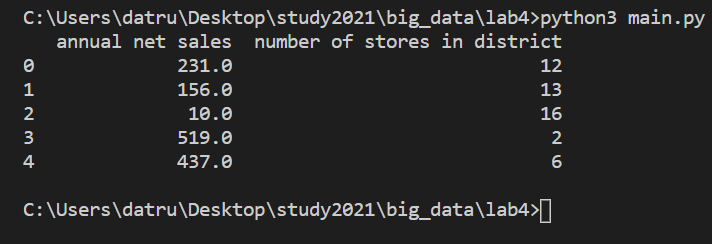




**Частина 2: Складання даних**

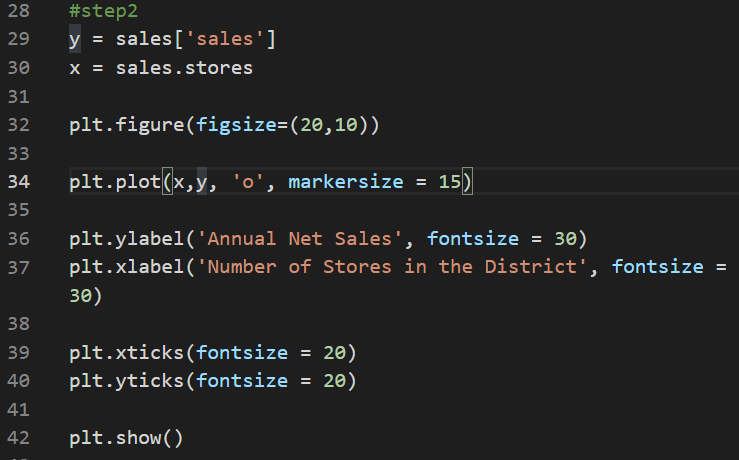
**Крок 1: Визначте співвідношення**

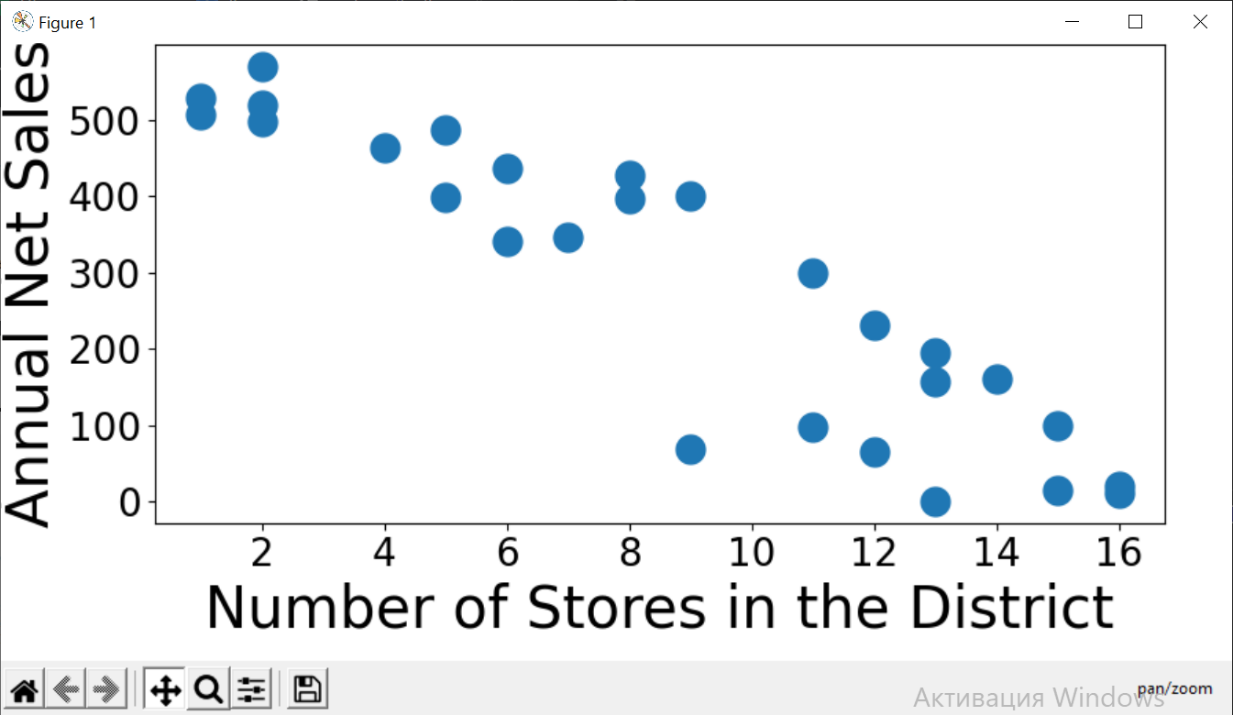
****

****

Кореляція є від’ємною

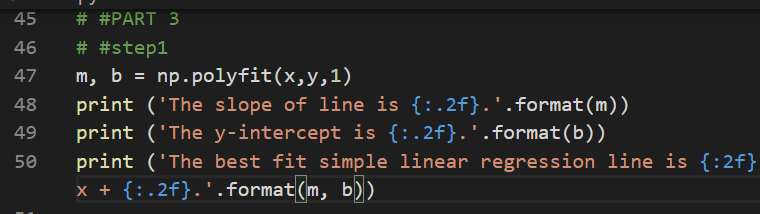
**Крок 2: Створіть сюжет.**

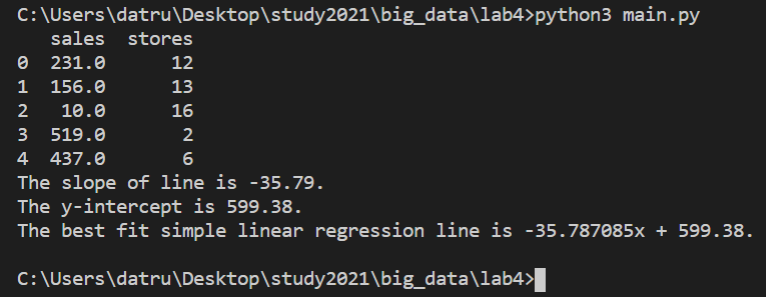


****

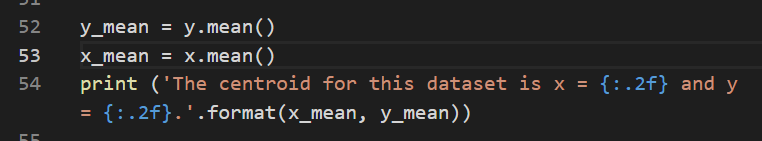
**Частина 3: Побудуйте просту лінійну регресію**

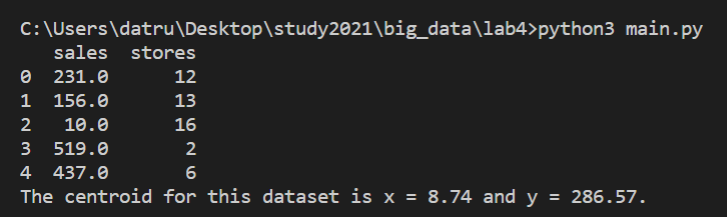
**Крок 1: Обчисліть нахил та перетин Y-лінії лінійної регресії.**

****

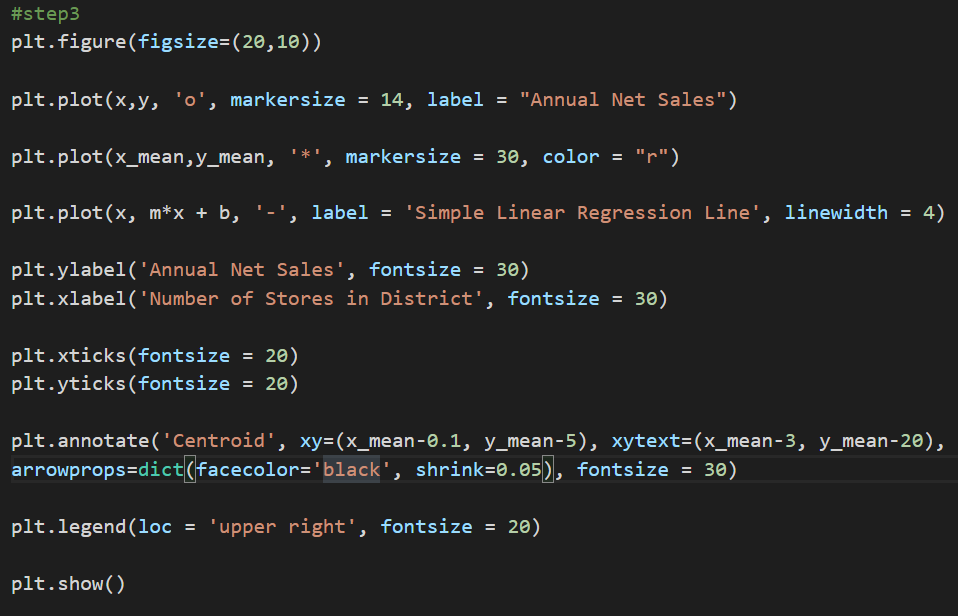
****

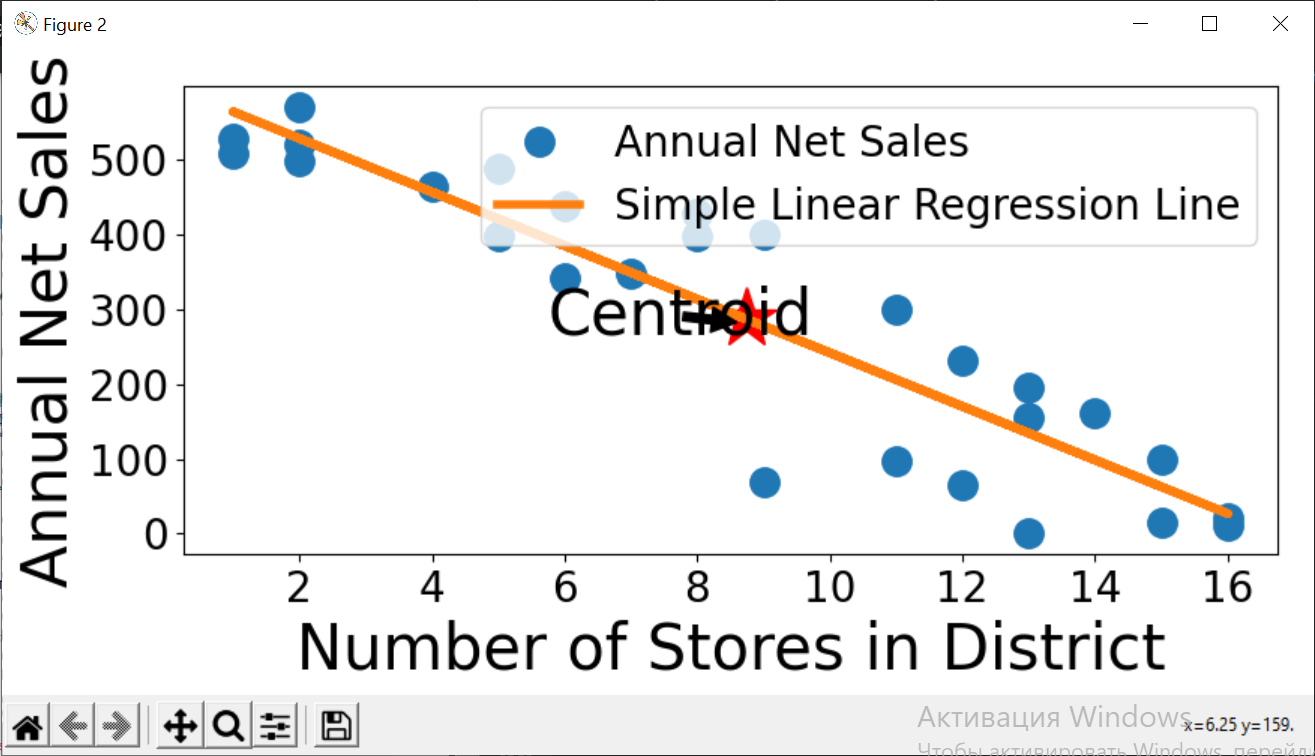
**Крок 2: Обчисліть центроїд**

****

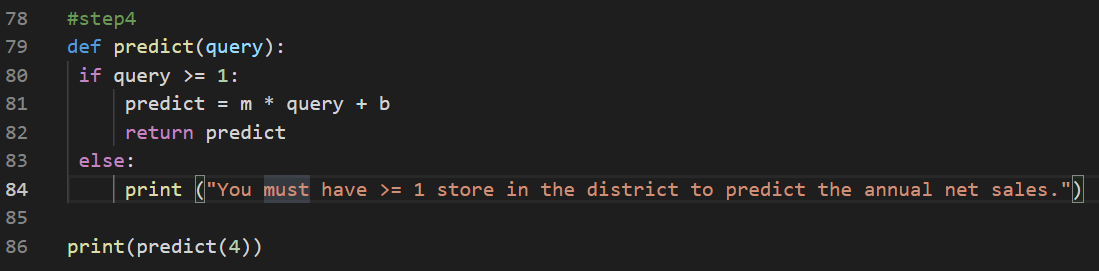
****

**Крок 3: Накладіть лінію регресії та центральну точку на графіку**

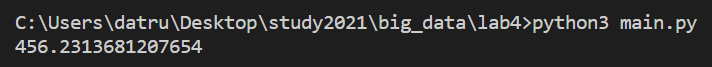
****

****

**Крок 4: Прогнозування**

****

Прогнозований чистий продаж з 4-х магазинів:



**Висновки**

В ході виконанні лабораторної роботи були використані навички моделювання лінійної регресії на базі набору даних з обсягів продажів в районах з допомогою бібліотек для мови Python, а саме Numpy, Pandas, Matplotlib. Ці бібліотеки були використані для зручної побудови графіків регресії чистого обсягу даних.