**Report Lab**

**Activity 6**

**ผู้สอน**

รศ.ดร.อรฉัตร จิตต์โสภักตร์

**กลุ่ม** : OPหลุดเรท

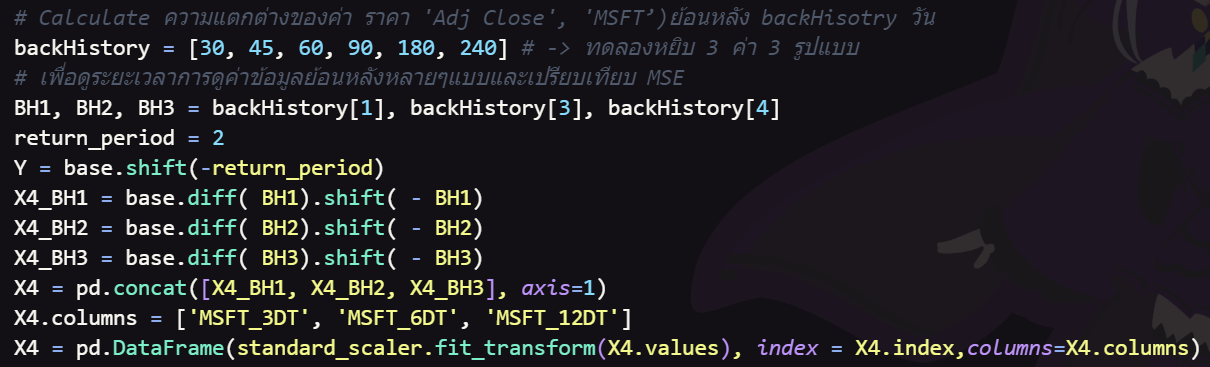
นายฐานพัฒน์ สิทธิพรชัยสกุล 63010256

นายณภัทร จิรารัตนกุลชัย 63010279

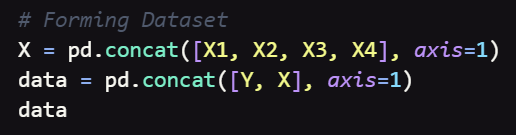
Data Preparation



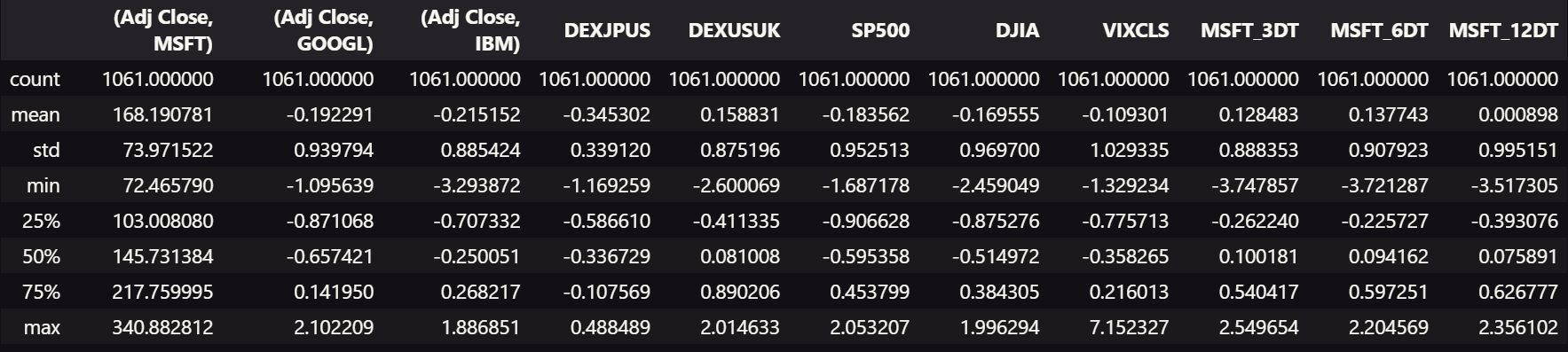
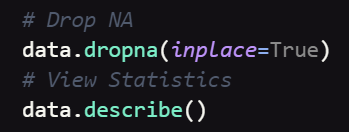
* เตรียมการข้อมูลโดยการดึงชุดข้อมูลตามที่อาจารย์กำหนด
* เลือกข้อมูลในคอลัมน์ที่ต้องการ
* ทำการ standardize ด้วย StandardScaler()



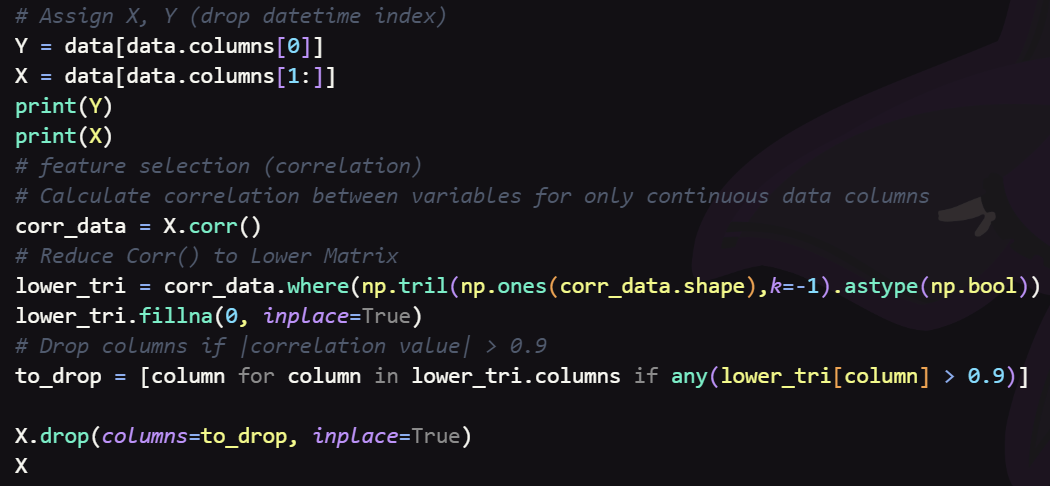
* ทำการหยิบค่ามา 3 ตัว สำหรับทำนายค่า
* จากนั้นนำมาทำ Dataframe



* รวม Dataframe X1 X2 X3 X4 เป็น Dataframe X
* นำ Dataframe X มารวมกับ Y



* ดรอปข้อมูลที่ใช้ไม่ได้ และตรวจสอบข้อมูล

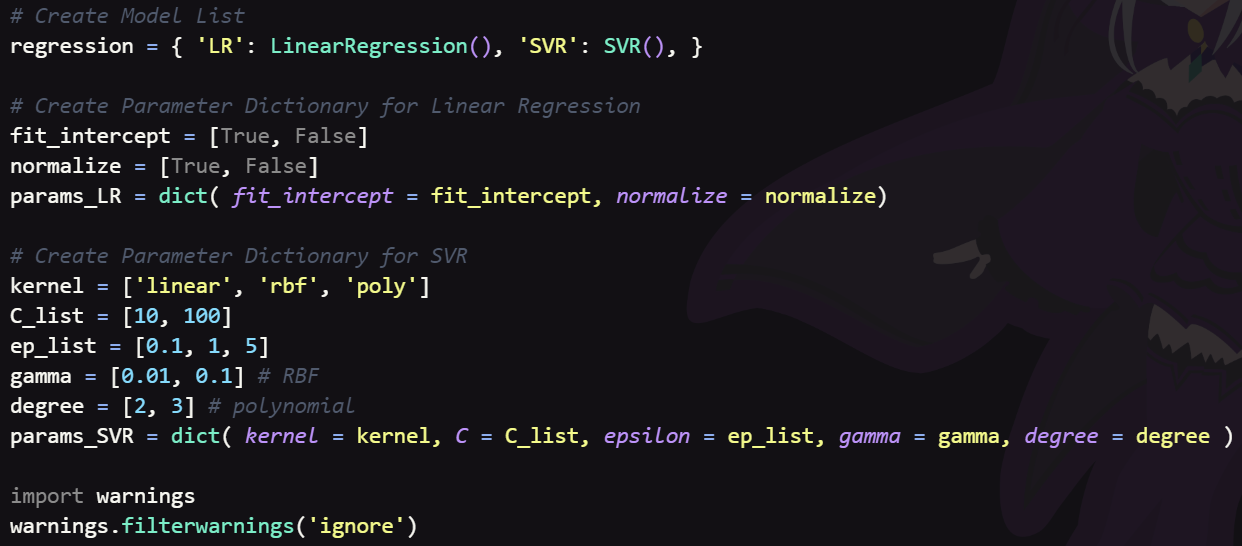


* แบ่งข้อมูลเป็น 2 ส่วนคือ X กับ Y
* จากนั้นนำ Dataframe X มา Correlation
* จากนั้นนำมาทำ Dimensional Reduction เพื่อดรอปค่าที่เกินในส่วนที่เราต้องการ

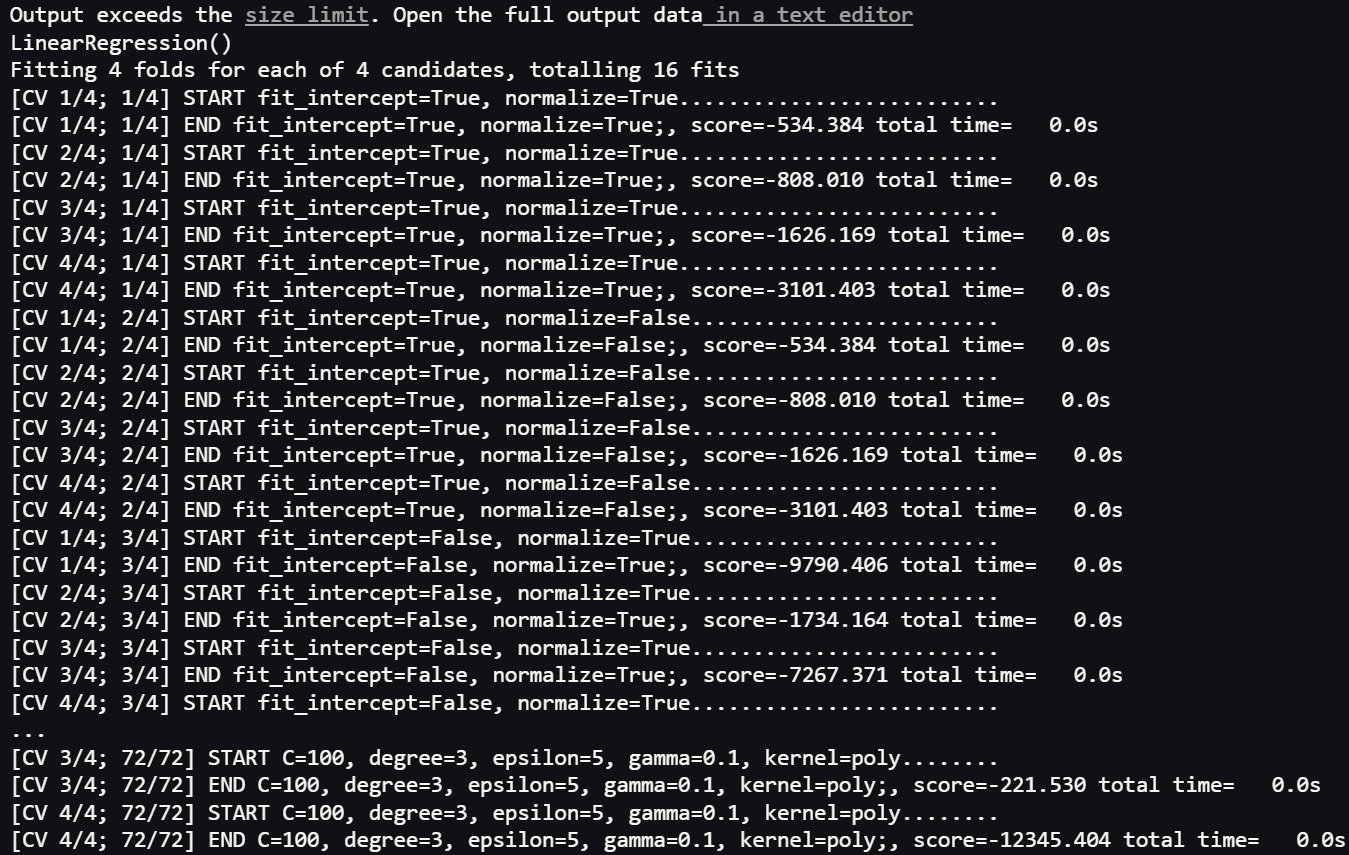


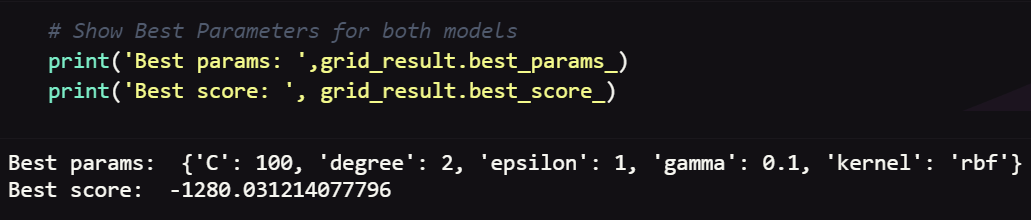
* ทำการแยกส่วนที่เอาไว้ Train กับ Test โดยแบ่งเป็น 2 options
* ทำ linear regression เพื่อหา prediction

GridSearchCV

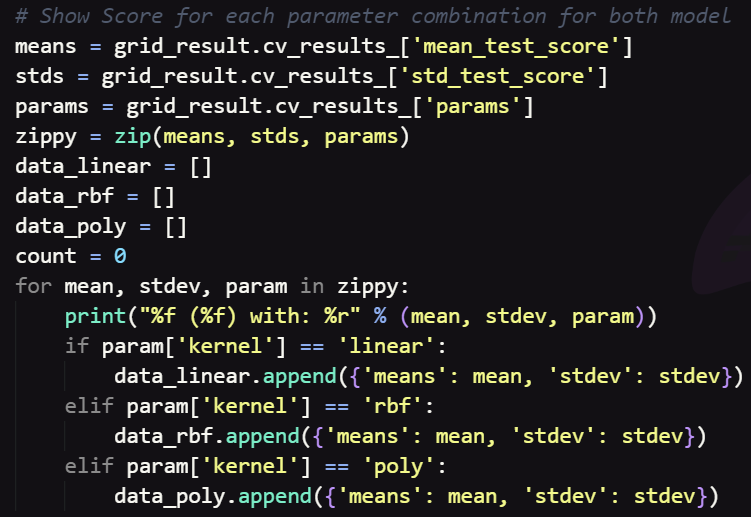


* สร้าง Model สำหรับทำ GridSearch
* สร้าง Parameter Dictionary สำหรับ Linear Regression
* สร้าง Parameter Dictionary สำหรับ SVR
* ส่วนด้านล่างมี warning เล็กน้อยซึ่งทำให้มองยาก กลุ่มผมเลยปิดไว้

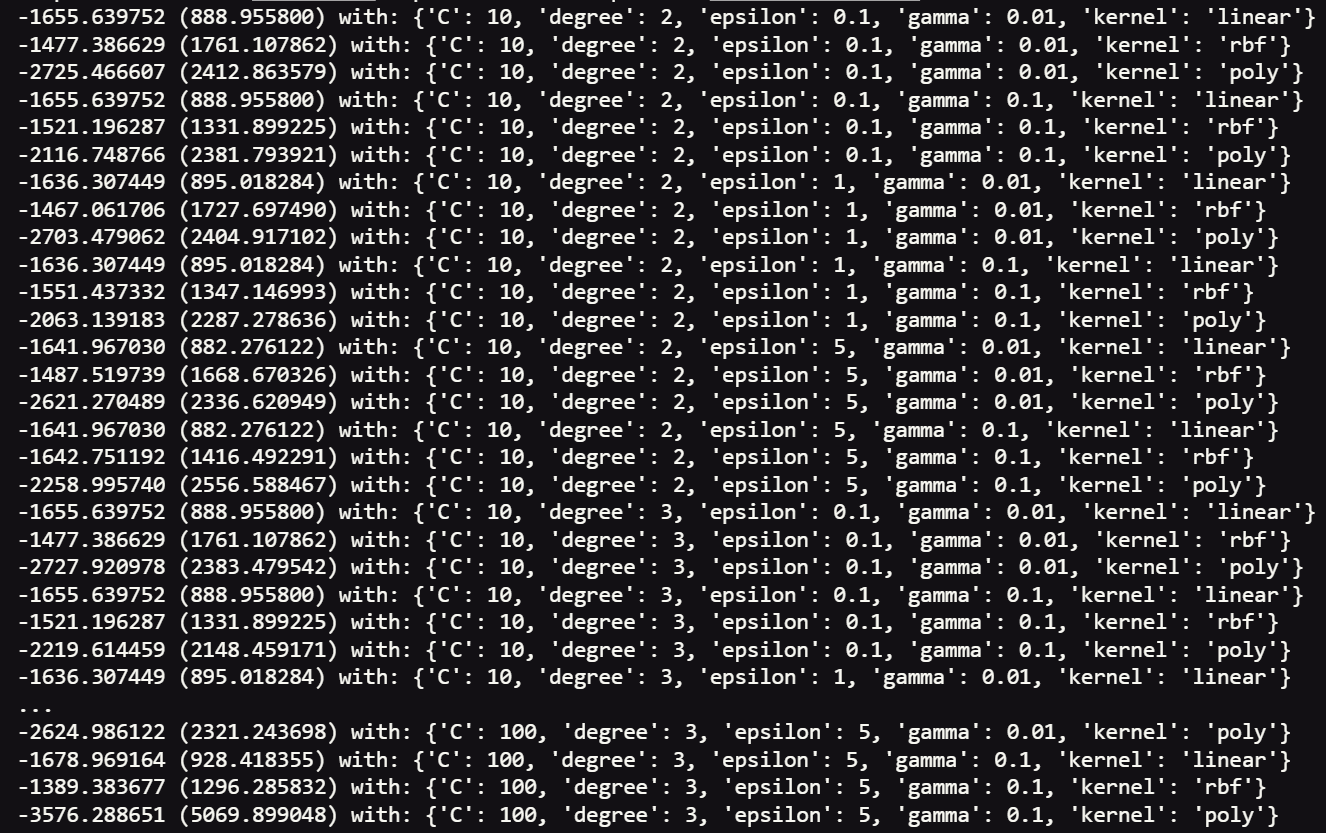




* ตรวจสอบ Parameter ที่ทำให้ผลลัพธ์ดีที่สุด

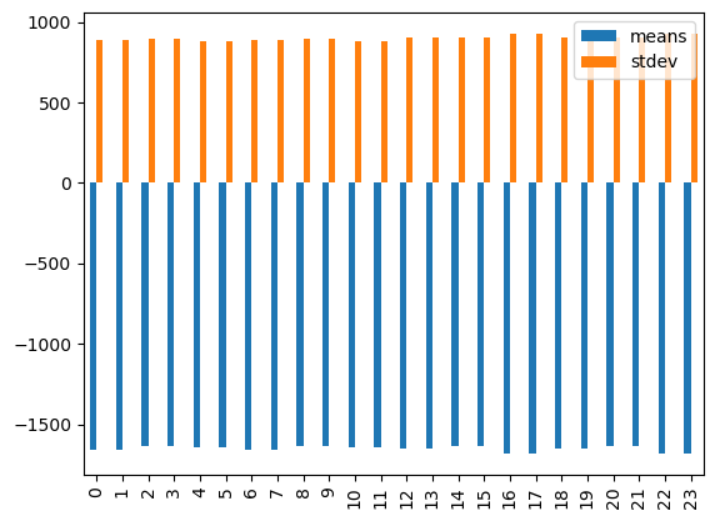
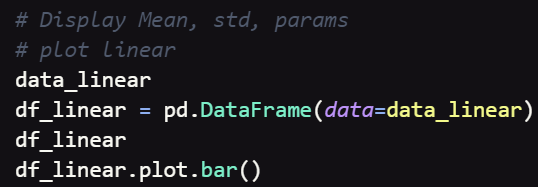


* ทำการแสดงผลค่า Score จากทุกๆ Parameter

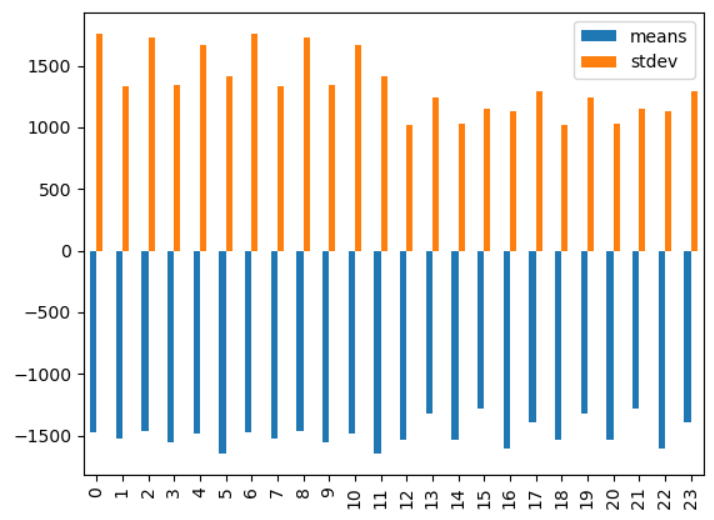
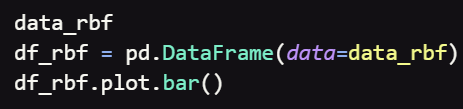


* การแสดงผลค่า Score จากหลายๆ Parameter

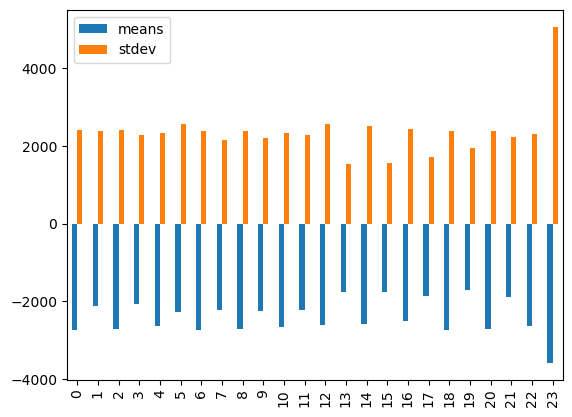
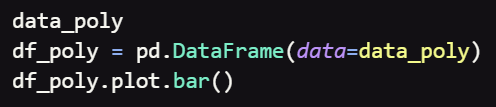
Option 1



* แสดงผลกราฟ Linear

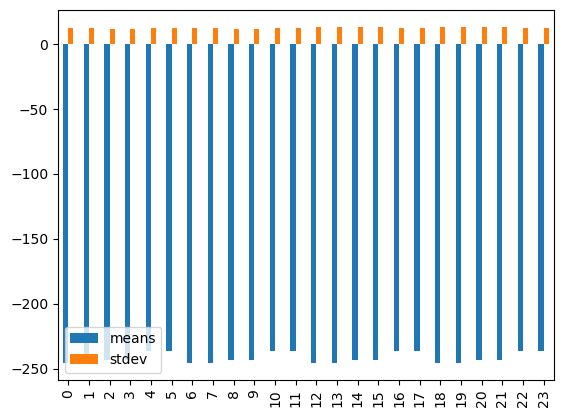
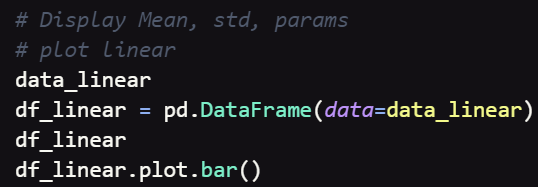


* แสดงผลกราฟ RBF

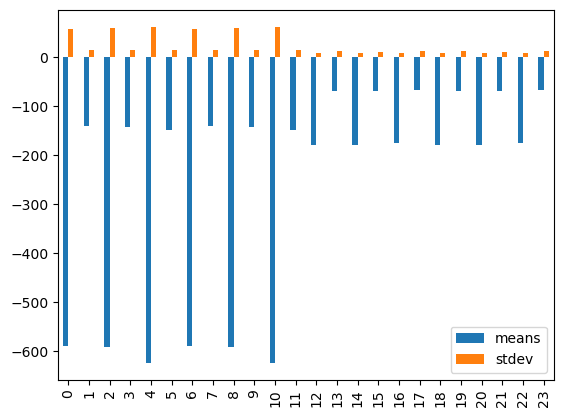
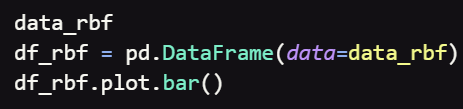


* แสดงผลกราฟ Poly

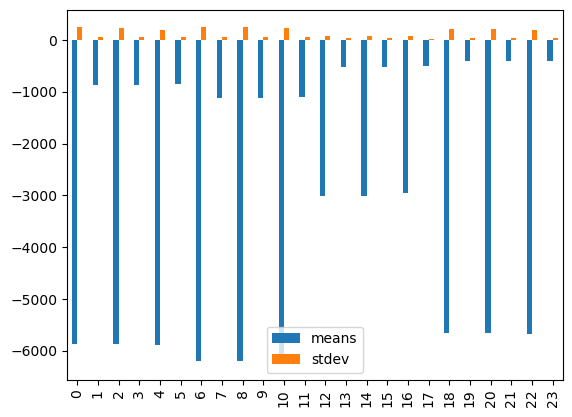
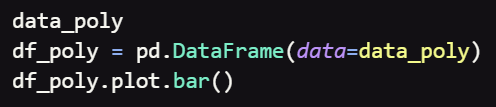
Option 2



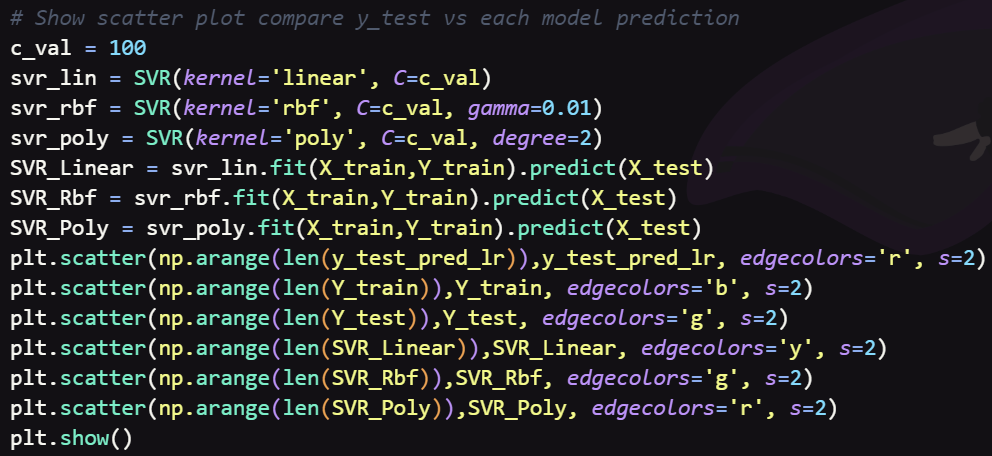
* แสดงผลกราฟ Linear



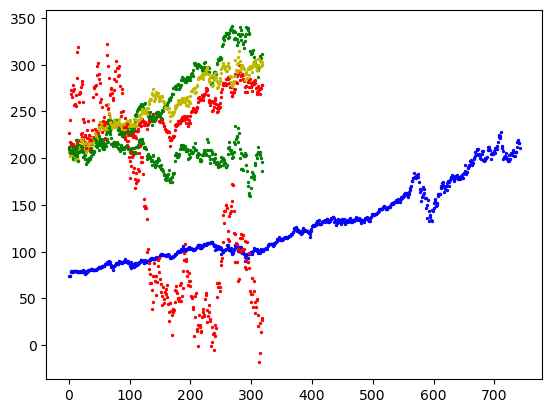
* แสดงผลกราฟ RBF

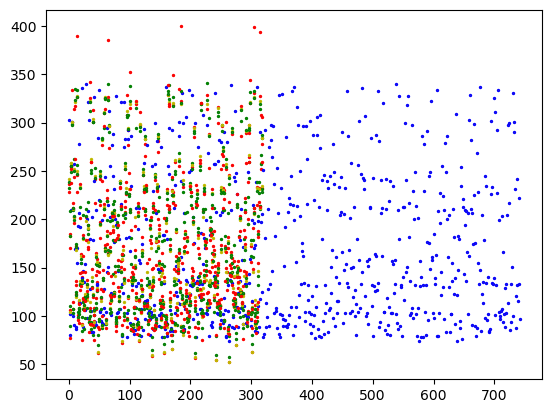


* แสดงผลกราฟ Poly



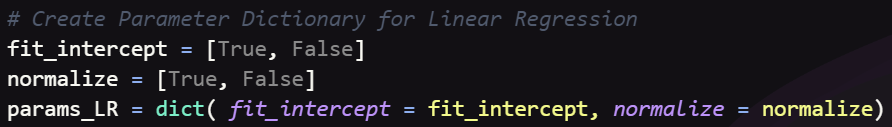
* แสดงผล scatter plot เพื่อเปรียบเทียบ y\_test กับค่าที่ทำนายทุกๆค่า

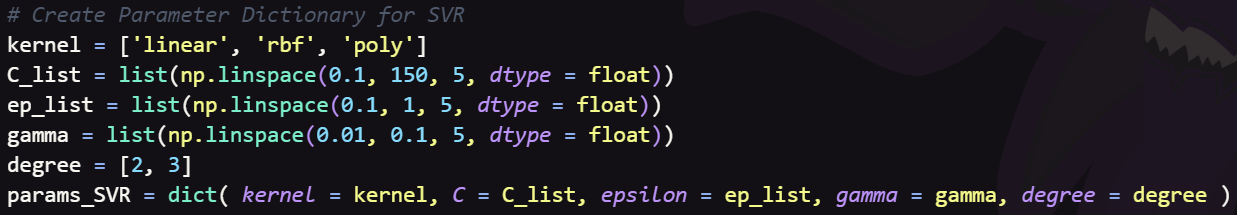
**option 1**

**option 2**

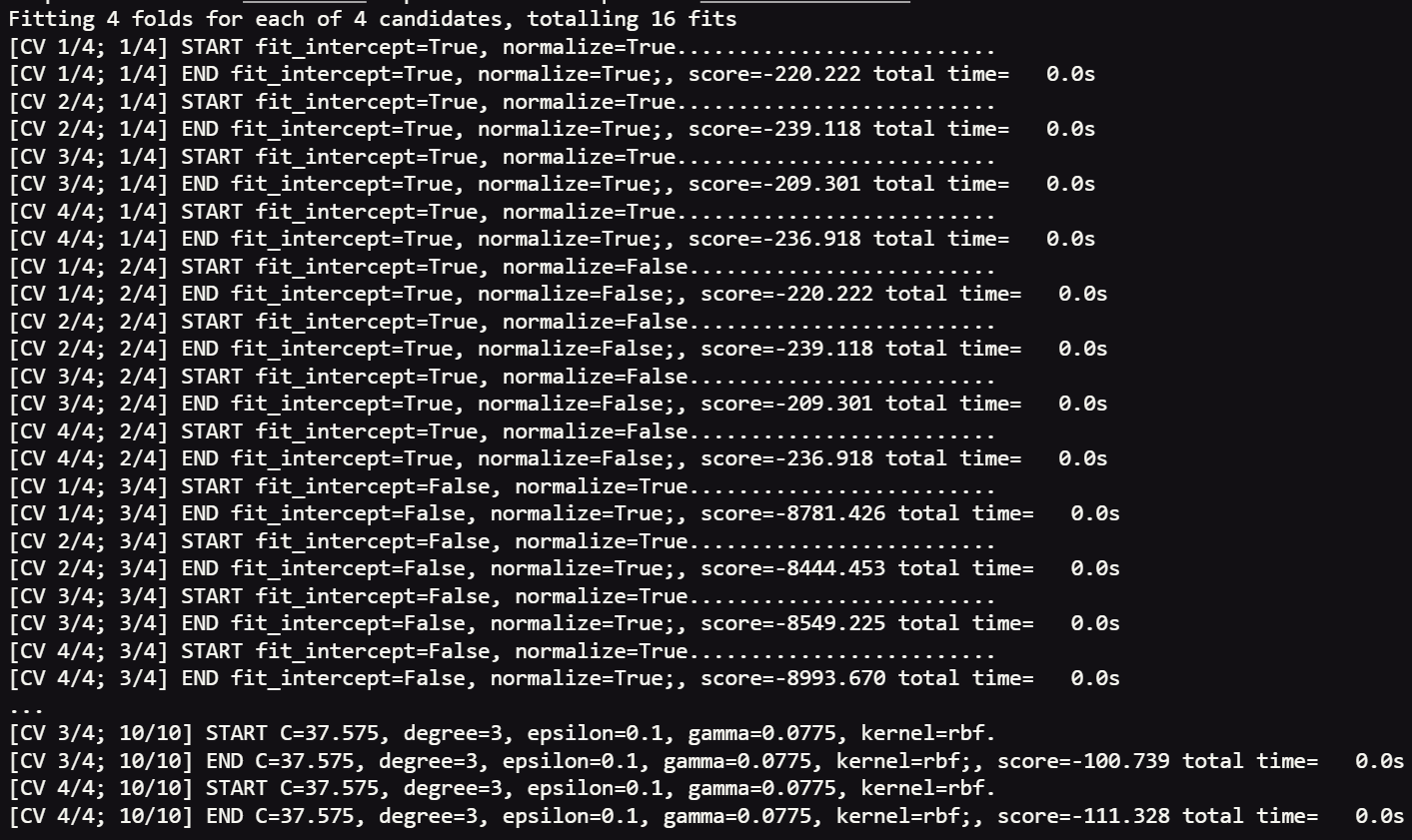
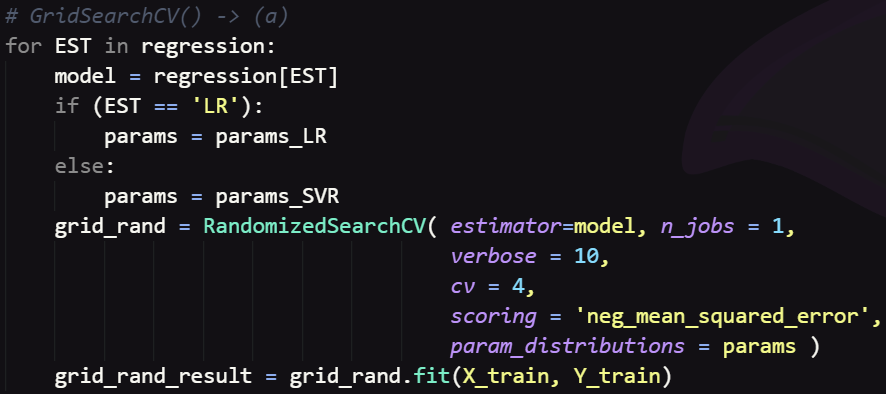
RandomizeSearchCV



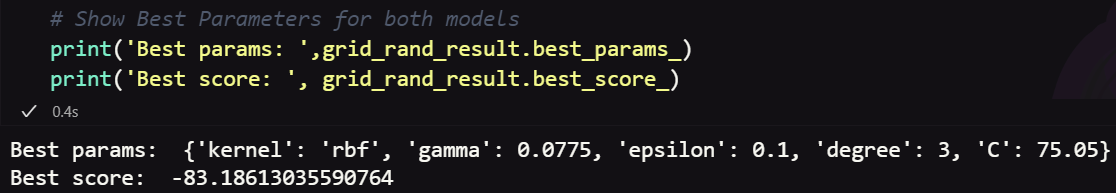




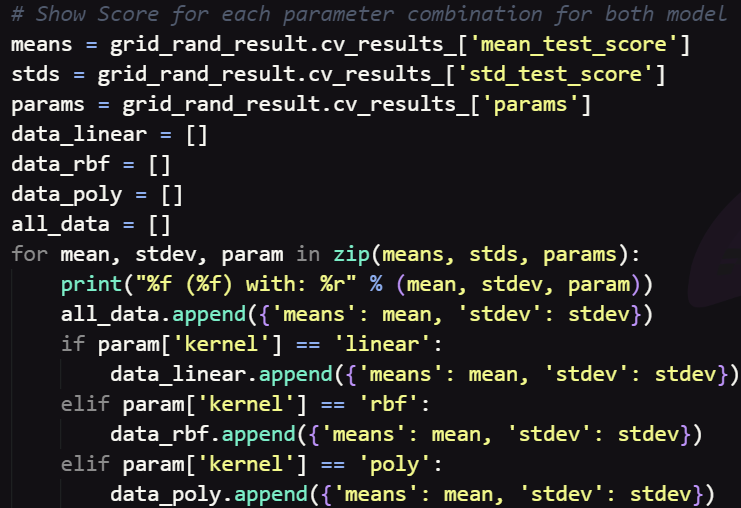
* สร้าง Model สำหรับทำ RandomizeSearch
* สร้าง Parameter Dictionary สำหรับ Linear Regression
* สร้าง Parameter Dictionary สำหรับ SVR



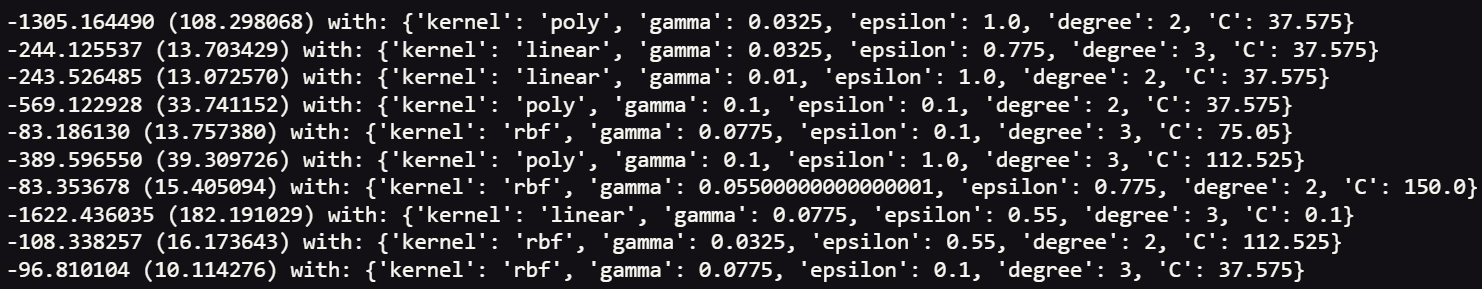
* การแสดงผลค่า Score จากหลายๆ Parameter



* ตรวจสอบ Parameter ที่ทำให้ผลลัพธ์ดีที่สุด

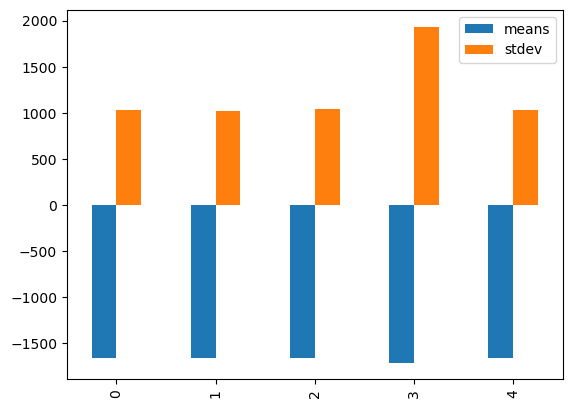
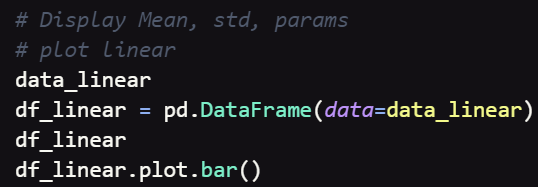


* ทำการแสดงผลค่า Score จากทุกๆ Parameter

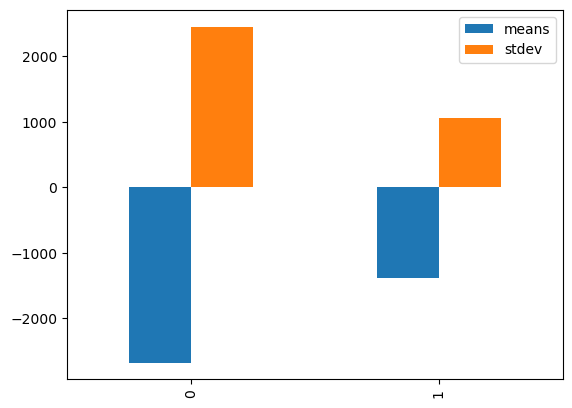
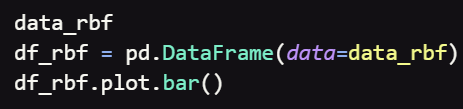


* การแสดงผลค่า Score จากหลายๆ Parameter

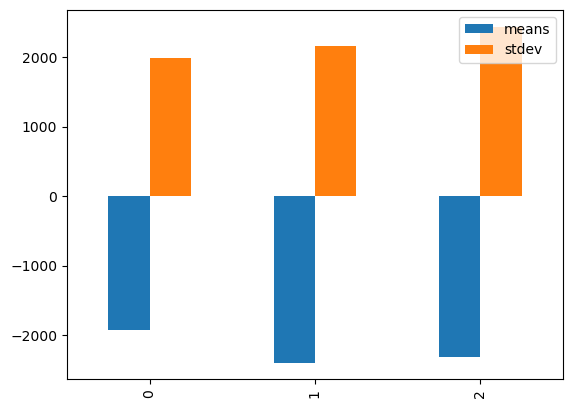
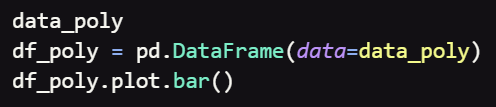
Option 1



* แสดงผลกราฟ Linear

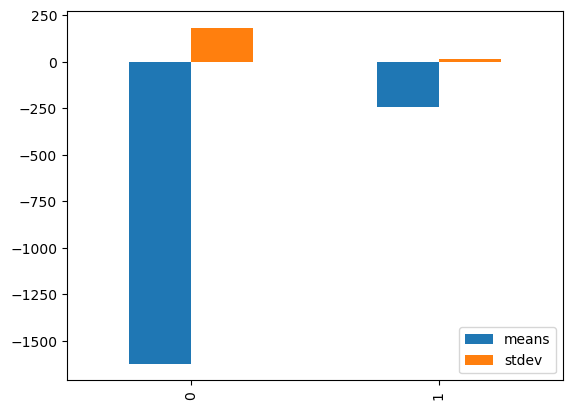
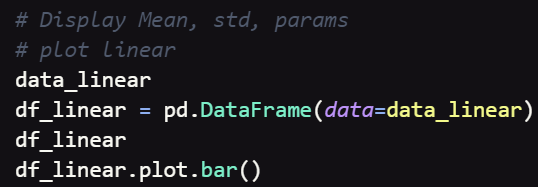


* แสดงผลกราฟ RBF

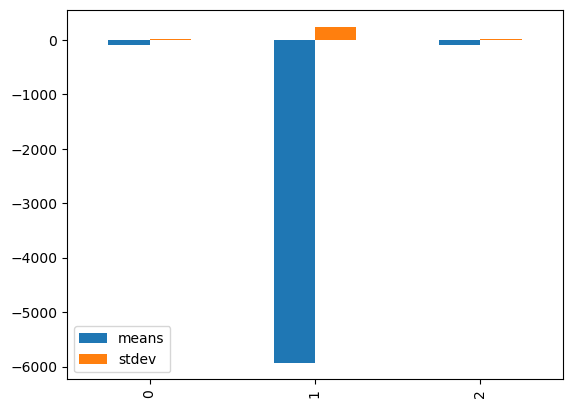
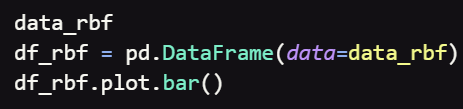


* แสดงผลกราฟ Poly

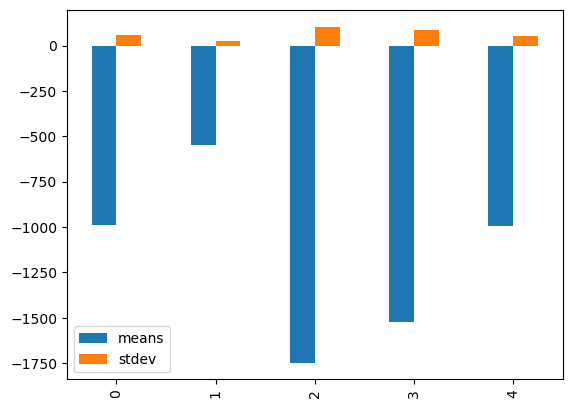
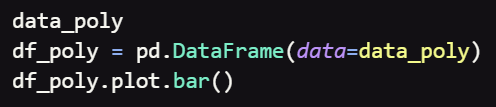
Option 2



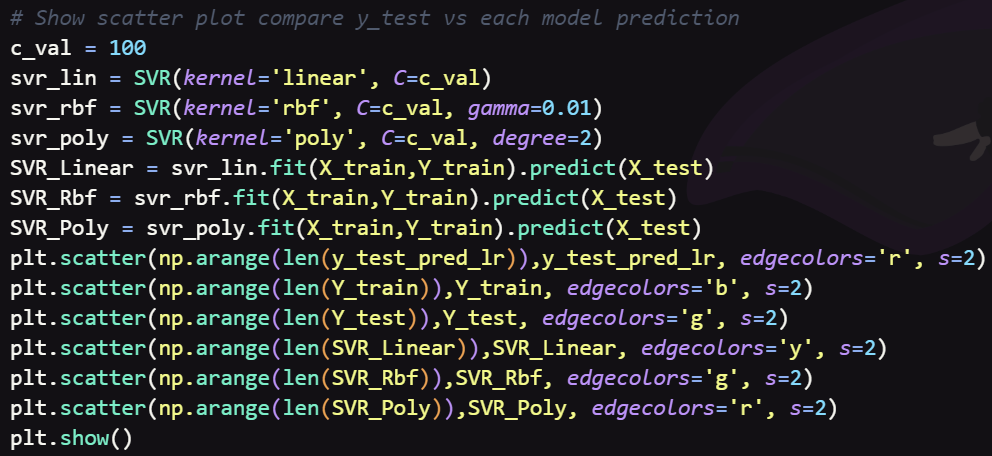
* แสดงผลกราฟ Linear



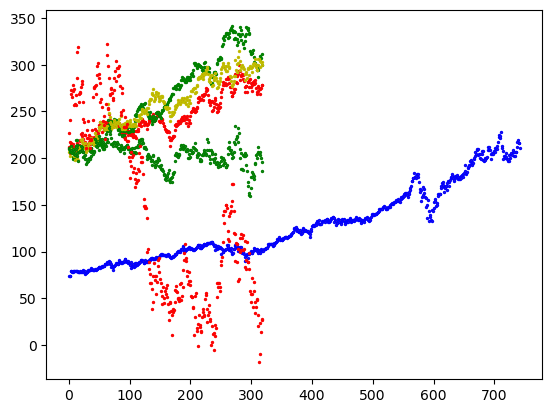
* แสดงผลกราฟ RBF

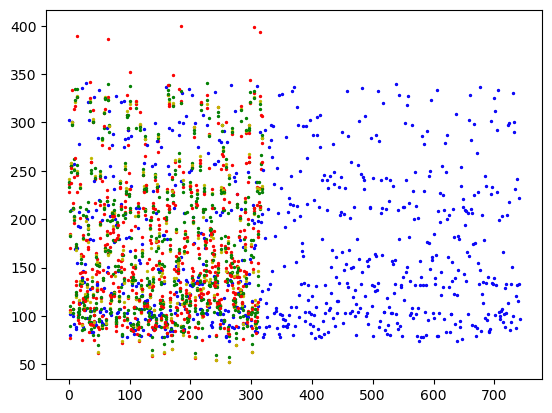


* แสดงผลกราฟ Poly



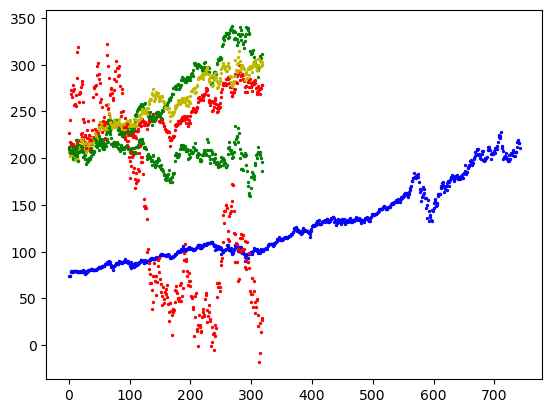
* แสดงผล scatter plot เพื่อเปรียบเทียบ y\_test กับค่าที่ทำนายทุกๆค่า

**option 1**

**option 2**

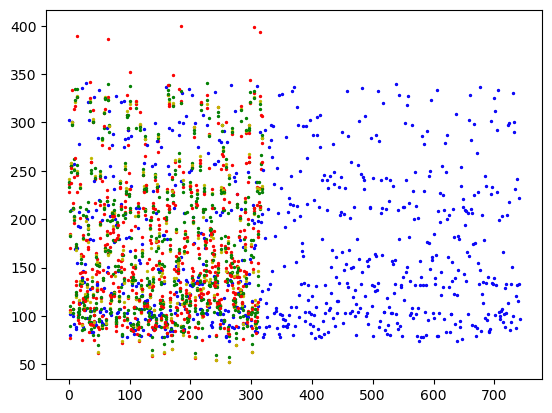
สรุปผลการทดลอง

จากการทดลองที่เราได้กำหนด option มาสองตัวจะสังเกตได้ว่าใน option 1 จะใช้วิธี train test split ซึ่งจะแบ่ง test size กับ train size ออกเป็นอัตราส่วน 3:7 โดยไม่ได้ shuffle จึงทำให้กราฟออกมาค่อนข้างเป็นอัตราส่วนกัน โดยจากภาพด้านล่างจุดสีฟ้าคือ y\_train ซึ่งแบ่งมาด้วยอัตราส่วน 0.7 จึงค่อนข้างเยอะกว่าข้อมูลที่เหลือ



**Option1**

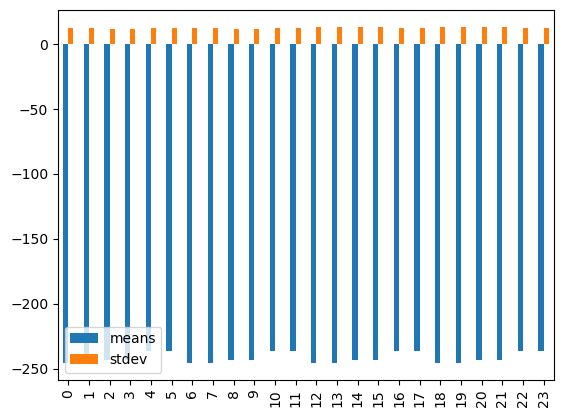
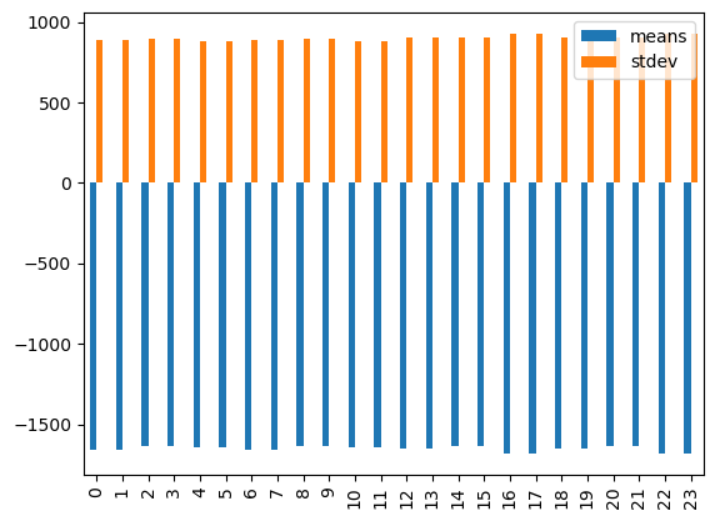
ส่วน option 2 จะใช้วิธี train test split แต่จะทำการ shuffle เพื่อให้ได้ test size กับ train size ออกมา โดยจะสังเกตได้ว่าข้อมูลมีการกระจายพอสมควร เพราะว่ามีการ shuffle ข้อมูลนั่นเอง



**Option2**

ส่วนความแตกต่างระหว่าง Grid Search และ Randomize Search ก็คือ Gird Search จะค้นหาค่าพารามิเตอร์จากชุดข้อมูลทั้งหมด ซึ่งอาจจะใช้เวลานานแต่ก็จะได้ค่าที่ดีที่สุดจากข้อมูลทั้งหมด

แต่ Randomize Search จะทำการสุ่มหาชุดข้อมูลมาแล้วทำการหาค่าพารามิเตอร์อีกทีซึ่งจะทำให้ค่าของชุดข้อมูลบางอันค่อนข้างที่จะไม่ใกล้เคียงกันโดยสังเกตได้จากกราฟด้านล่าง



สังเกตได้จากกราฟด้านซ้ายซึ่งเป็นการใช้ Grid Search ส่วนภาพด้านขวาจะเป็นการใช้ Randomize Search