

# ข้อมูลดิบ Raw Data & Normalized Data

## ข้อมูลดิบก่อนทำ Normalisation คือ

1. ค่าข้อมูลที่มีช่วงกว้างต่างกันมาก วิเคราะห์ร่วมกันทำได้ยากเพราะข้อมูลมีสเกลต่างกันมาก
2. ส่งผลต่อประสิทธิภาพของโมเดล หากข้อมูลมีสเกลไม่เท่ากัน เพราะทำให้โมเดลเรียนรู้ช้าหรือเกิด bias ได้
3. ข้อมูลดิบที่มีช่วงค่าต่างกันมาก ทำให้การเปรียบเทียบแนวโน้มหรือการวิเคราะห์เชิงสถิติเป็นไปได้ยาก
4. อาจเกิดปัญหาข้อมูล outliers หรือค่าผิดปกติ
5. ลดความแม่นยำในการวิเคราะห์หลายมิติ หากข้อมูลมีสเกลต่างกันอย่างมาก การรวมข้อมูลหลายตัวเข้าด้วยกันเพื่อวิเคราะห์ร่วมกันจะทำให้ผลลัพธ์เบี่ยงเบน

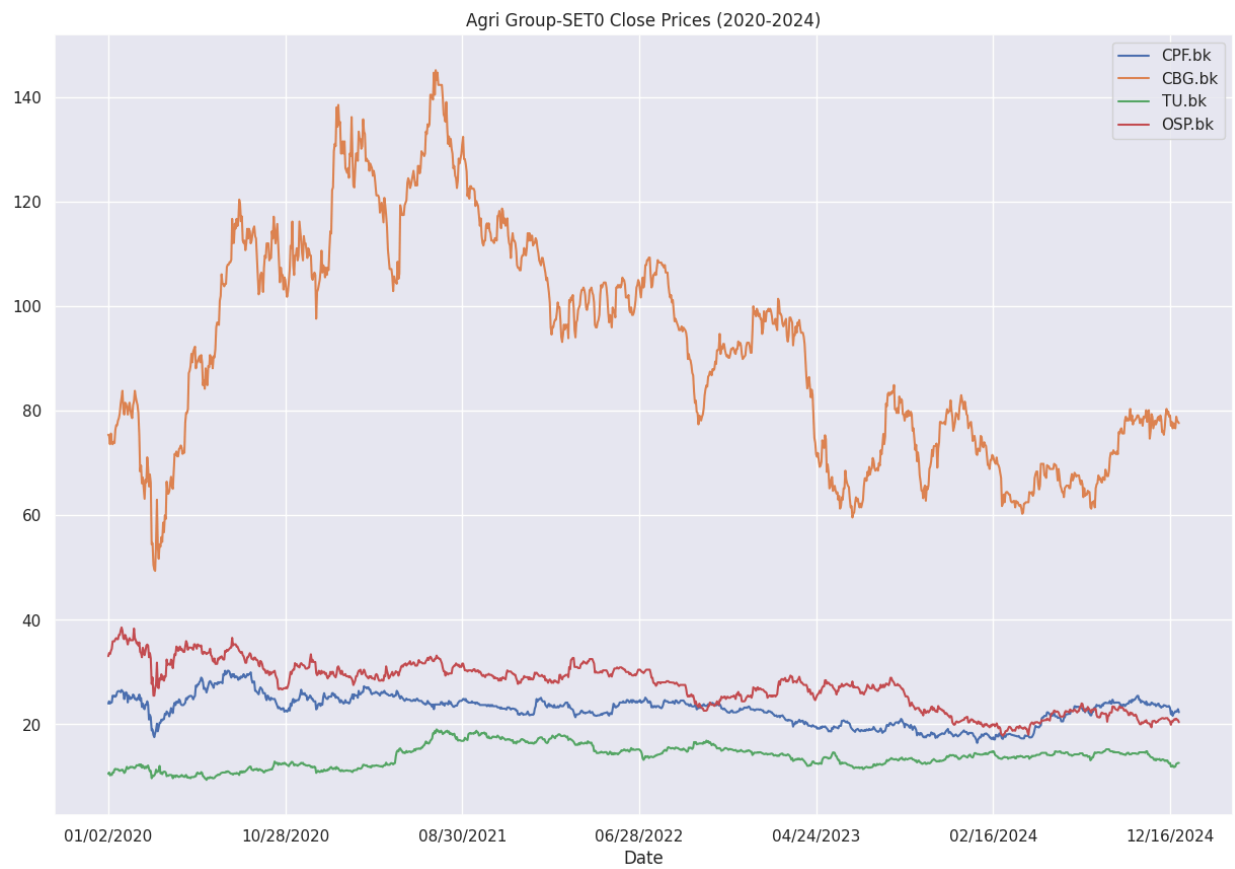
## ข้อมูลหลังการทำ Normalization

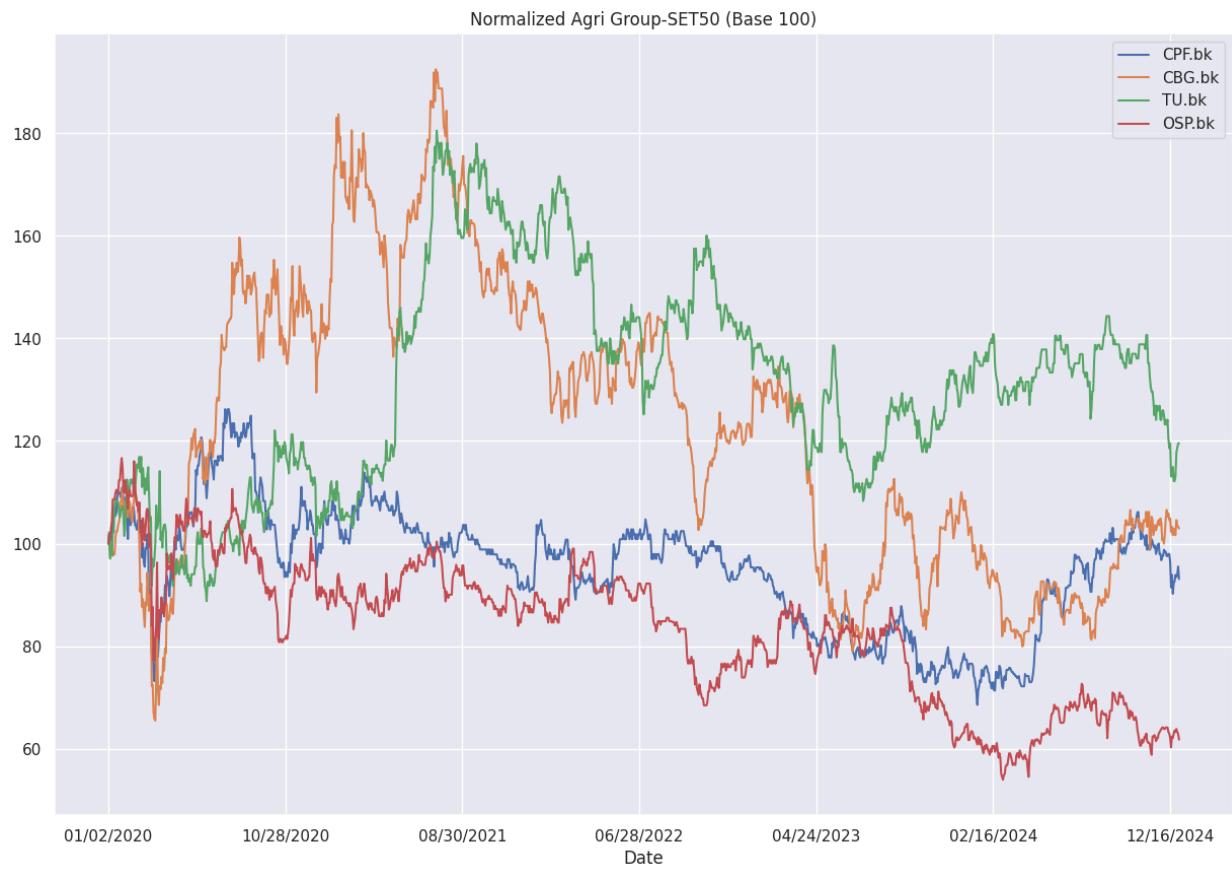
การทำ Normalization จะต้องนำราคาหุ้นแต่ละตัวมา ปรับสเกลให้อยู่บนสเกลเดียวกัน เพื่อให้สามารถเปรียบเทียบการเปลี่ยนแปลงของราคาหุ้นหลายตัวพร้อมกันได้อย่างเหมาะสม

หลักการทำ Normalization ที่ใช้คือการหารราคาหุ้นแต่ละวันด้วยราคาหุ้นวันแรก (`stocks / stocks.iloc[0]`) จากนั้นคูณด้วย 100 เพื่อกำหนดให้ราคาวันแรกเริ่มต้นที่ 100 เท่ากันสำหรับทุกหุ้น

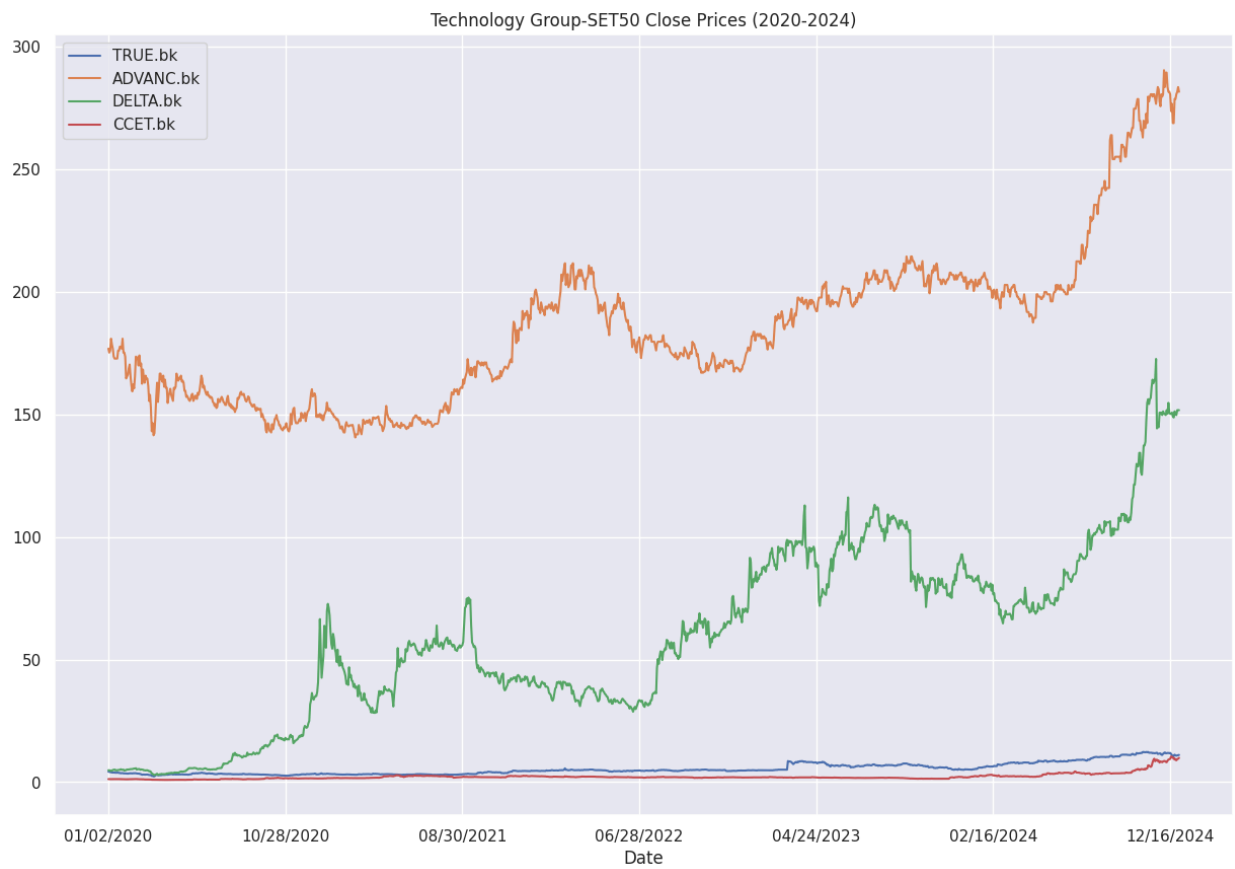
ผลลัพธ์คือ หุ้นทุกตัวมีจุดเริ่มต้นเดียวกัน และการเคลื่อนไหวของแต่ละหุ้นหลังจากนั้นจะอยู่บนสเกลเดียวกัน ทำให้สามารถพล็อตกราฟและเปรียบเทียบการเปลี่ยนแปลงราคาระหว่างหุ้นต่าง ๆ ได้อย่างชัดเจน และวิเคราะห์แนวโน้ม หรือความสัมพันธ์ของผลตอบแทนแต่ละหุ้นบนสเกลเดียวกันได้ง่ายขึ้น

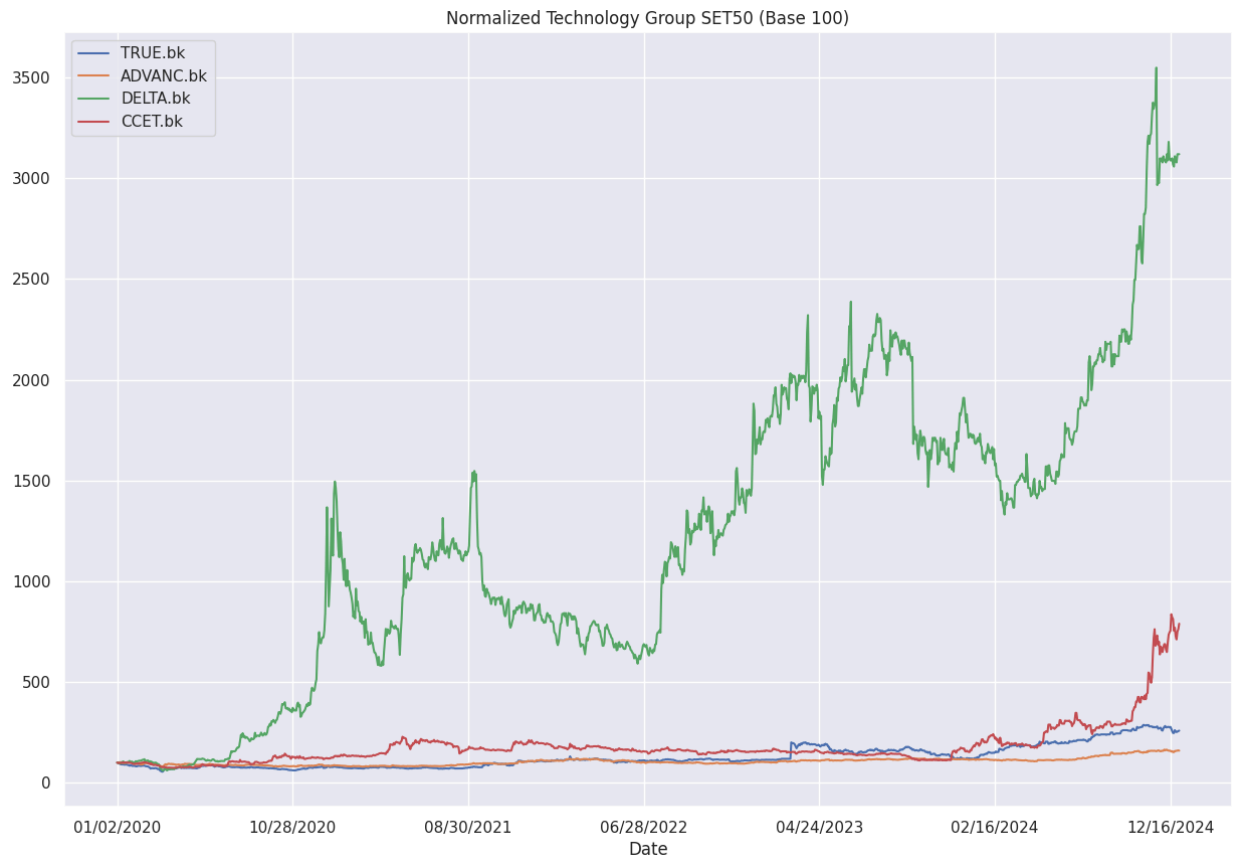
## เกษตรและอุตสาหกรรมอาหาร





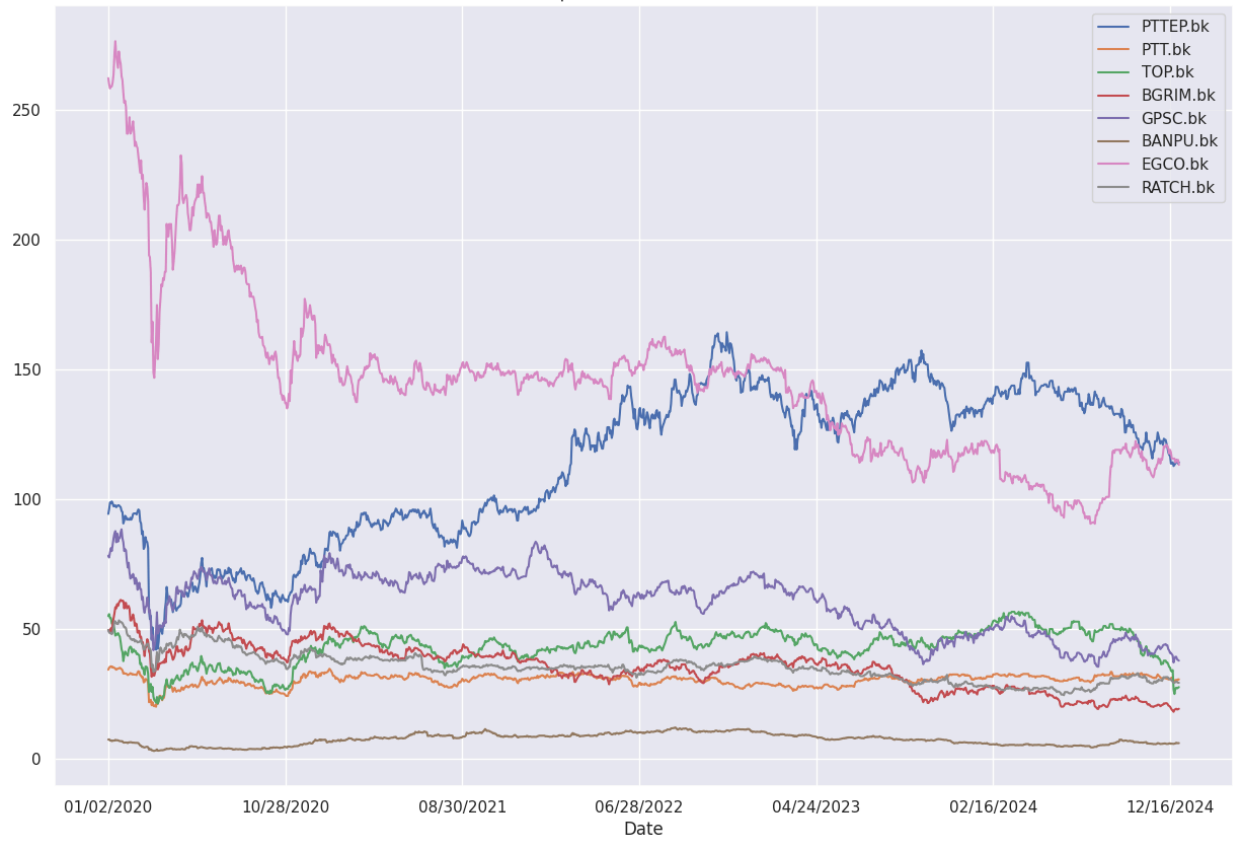
เทคโนโลยี

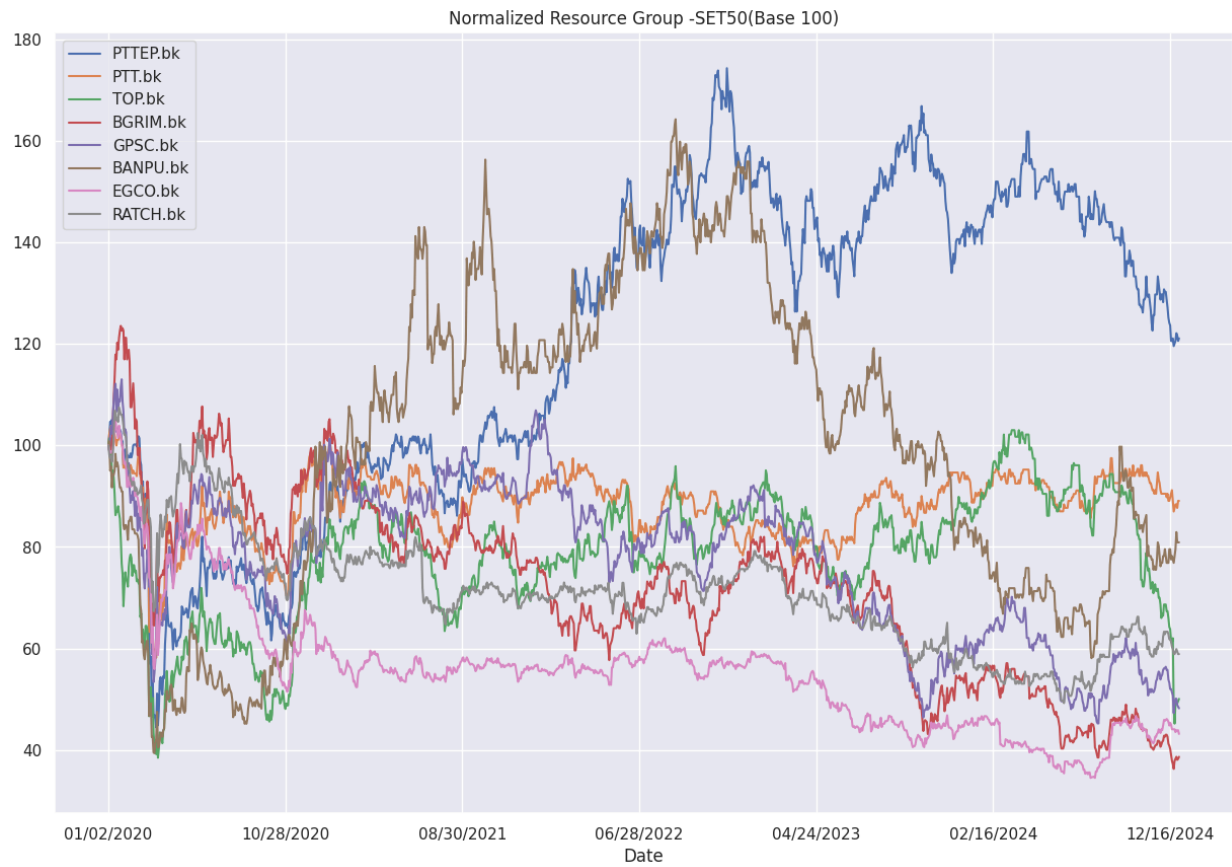




ทรัพยากร

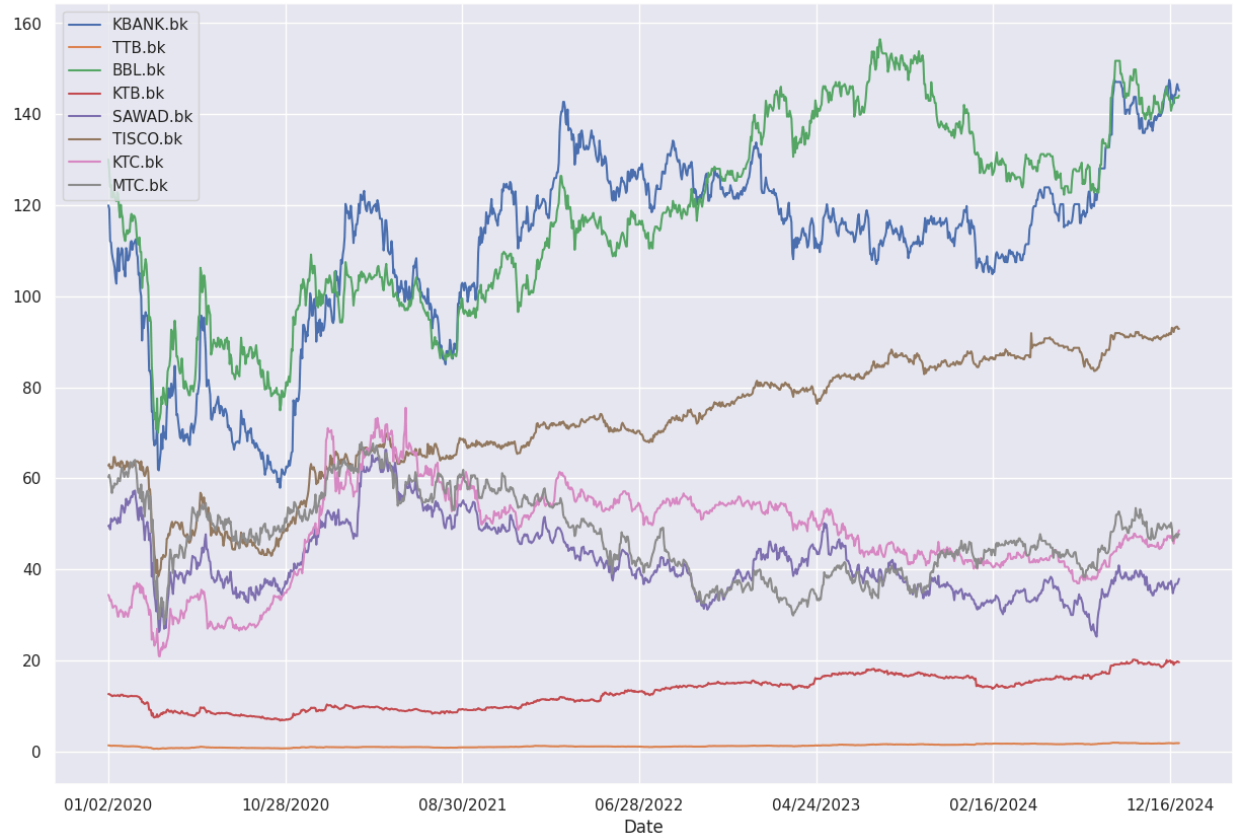
Resource Group -SET50 Close Prices (2020-2024)



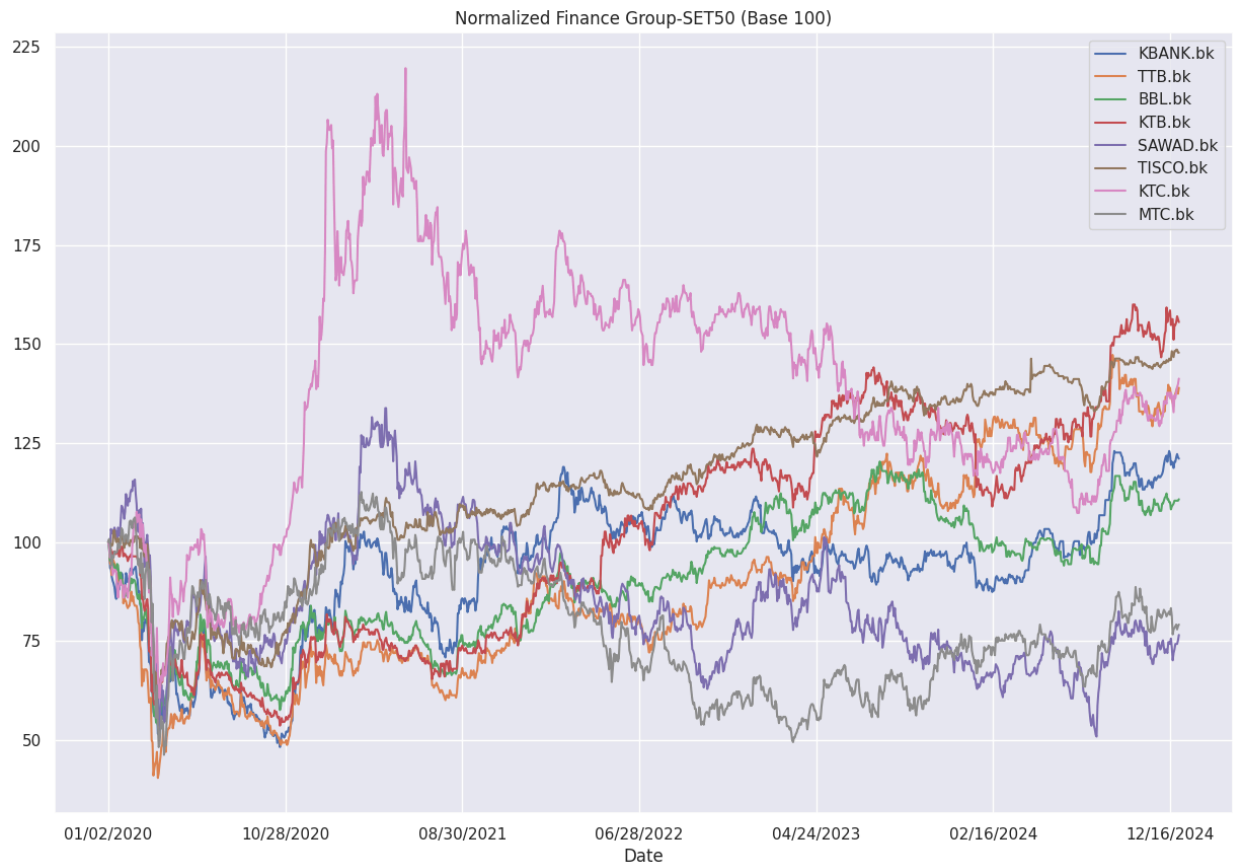


## ธุรกิจการเงิน

Finance Group-SET50 Close Prices (2020-2024)

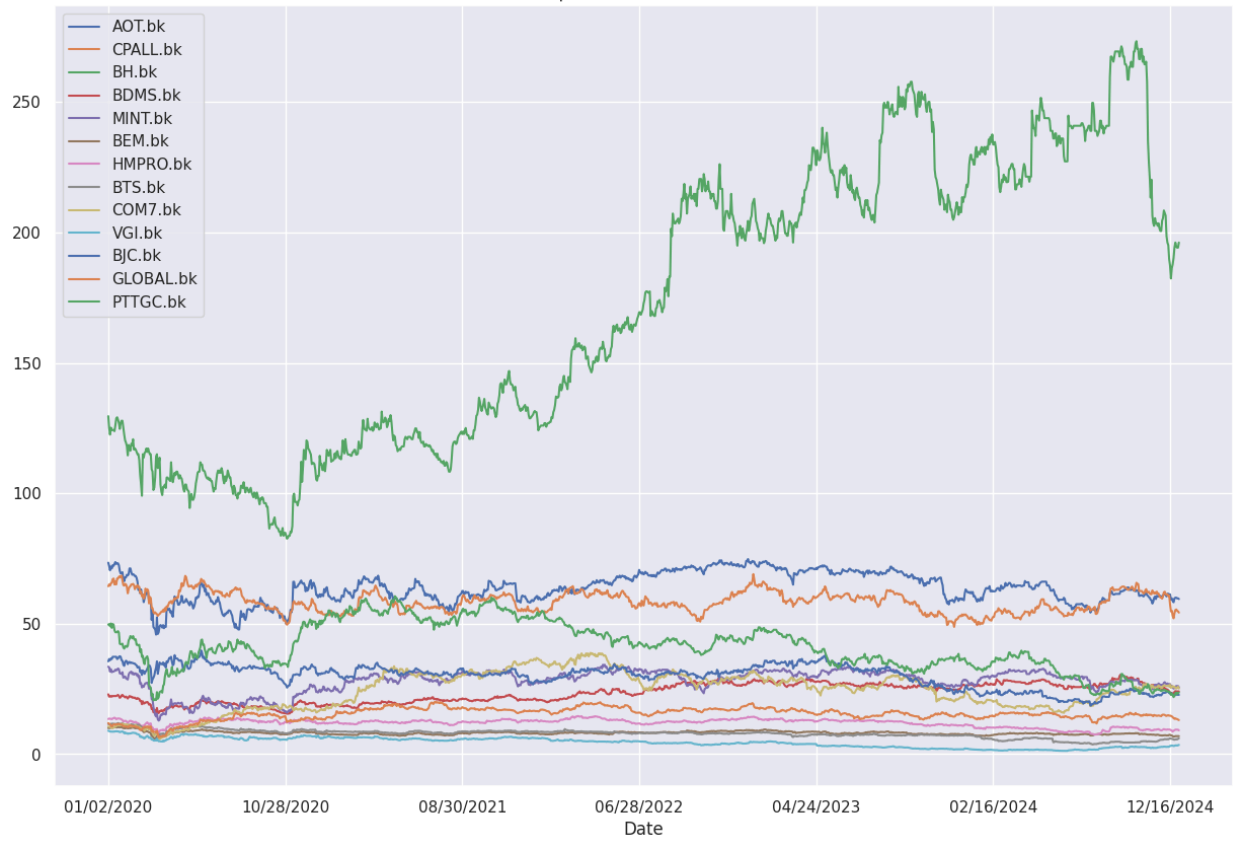


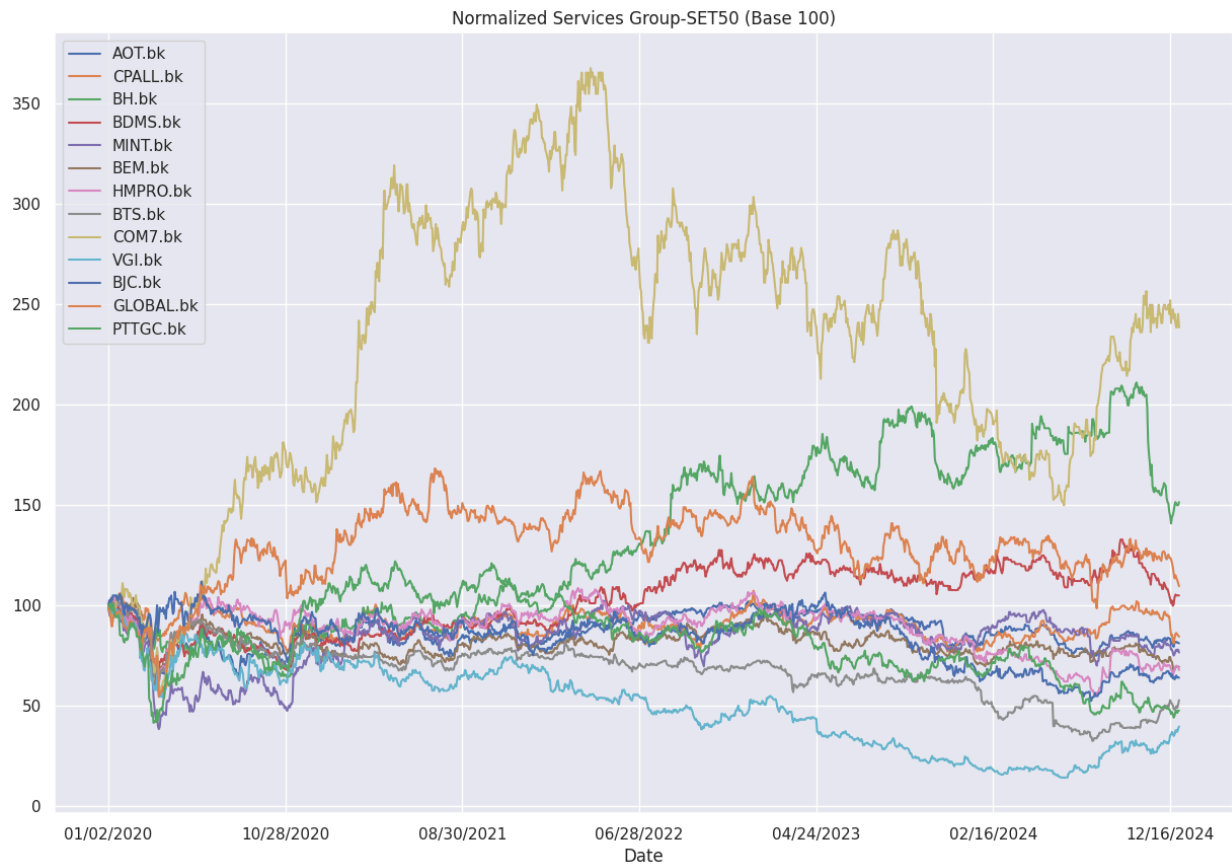




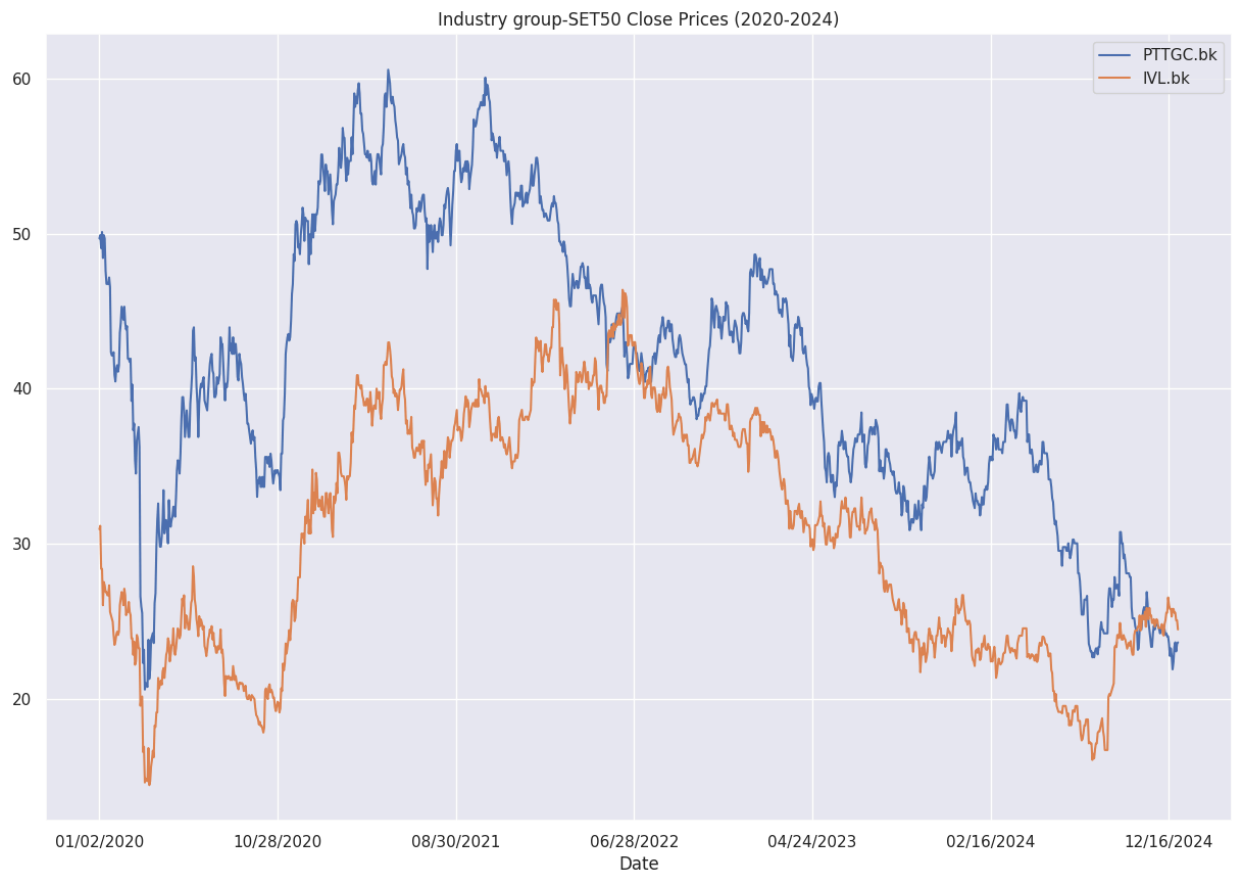
**บริการ**

Services Group-SET50 Close Prices (2020-2024)



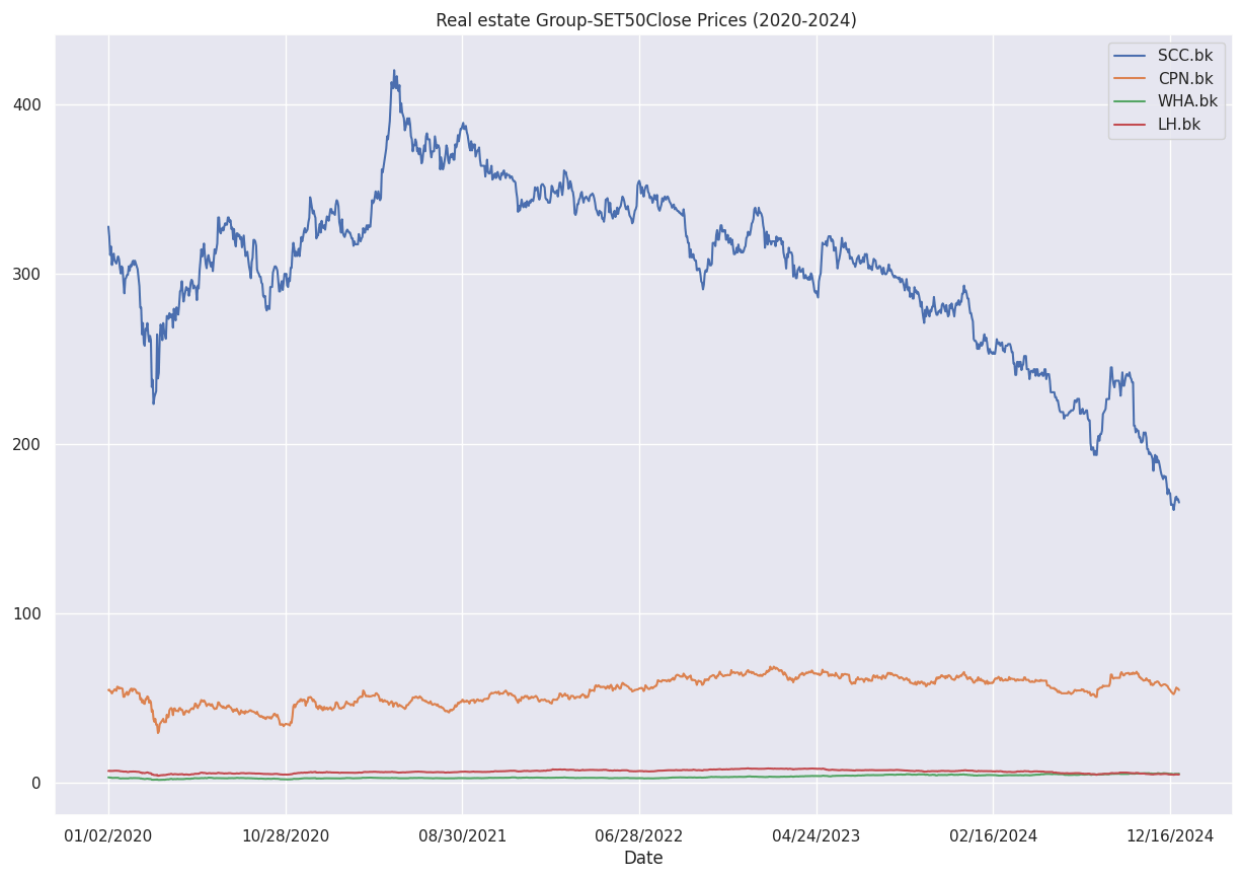


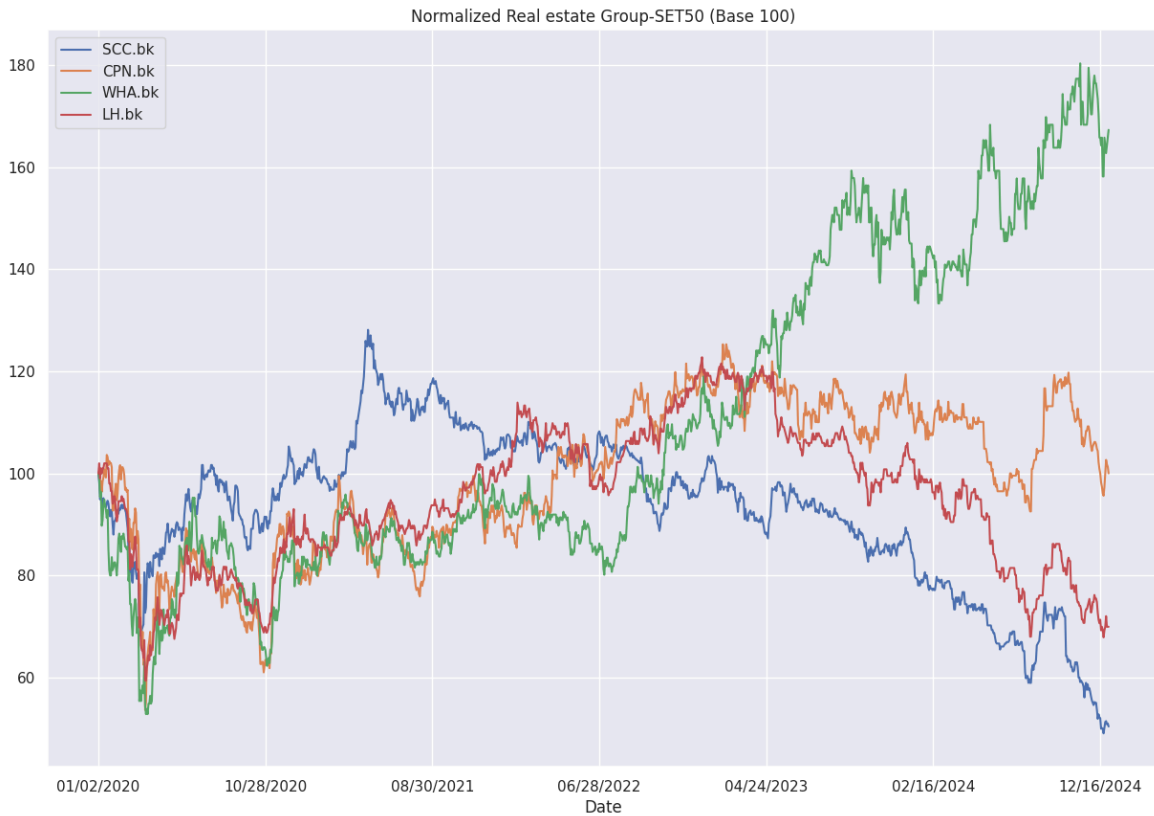
## สินค้าอุตสาหกรรม





## อสังหาริมทรัพย์และก่อสร้าง





## ข้อมูลที่น่าสนใจ

พล็อตราคาดีบของหุ้นในกลุ่มเกษตรและอุตสาหกรรมอาหาร ตั้งแต่ปี 2020 ถึง 2024 โดยยกเว้นหุ้น 'ITC.bk' เนื่องจากหุ้นตัวนี้ไม่มีข้อมูลราคาภายในช่วงเวลาที่กำหนด

หุ้นในกลุ่มนี้ที่อยู่ใน SET50 และอยู่ในช่วงเวลา 2020-2024 มีทั้งหมด 4 บริษัท

ชื่อ	กลุ่มอุตสาหกรรม	หมวดธุรกิจ
บริษัท เจริญโภคภัณฑ์อาหาร จำกัด (มหาชน)	เกษตรและอุตสาหกรรมอาหาร	อาหารและเครื่องดื่ม
บริษัท คาราบาวกรุ๊ป จำกัด (มหาชน)	เกษตรและอุตสาหกรรมอาหาร	อาหารและเครื่องดื่ม
บริษัท ไทยยูเนี่ยน กรุ๊ป จำกัด (มหาชน)	เกษตรและอุตสาหกรรมอาหาร	อาหารและเครื่องดื่ม
บริษัท โอสโสภา จำกัด (มหาชน)	เกษตรและอุตสาหกรรมอาหาร	อาหารและเครื่องดื่ม

---

## หลังจากที่เราทำการ Normalization เรียบร้อยแล้ว ขั้นตอนต่อไปคือการคำนวณผลตอบแทนรายวัน (Daily Return) ในรูปแบบ Log Return

คำนวณ Log Return เพื่อการเปลี่ยนแปลงของราคาหุ้นในแต่ละวัน

วิธีการคำนวณ Log Return คือเอาราคาของหุ้นวันนี้หารด้วยราคาของหุ้นวันก่อนหน้า แล้วนำมาลอการิทึมจากนั้นลบแถวแรกที่เป็นค่าว่างออกไป (dropna()) เราก็จะได้ผลตอบแทนรายวันของหุ้นแต่ละตัวในรูปแบบ Log Return ที่พร้อมใช้สำหรับการวิเคราะห์ต่อไป

สูตรคำนวณ Log Return:

python

```
log_return = np.log(normStocks / normStocks.shift(1))
```

- ♦ `normStocks.shift(1)` หมายถึงราคาหุ้นของวันก่อนหน้า
- ♦ นำราคาของหุ้นวันนี้หารด้วยราคาของวันก่อน จากนั้นนำมาลอการิทึมธรรมชาติ (`np.log()`) เพื่อให้ได้อัตราผลตอบแทน

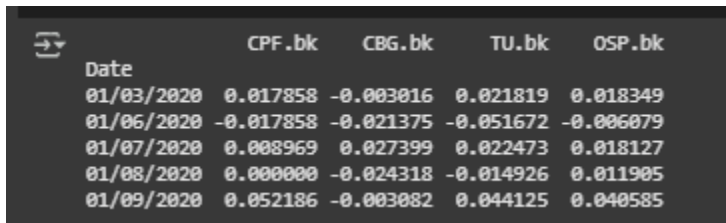
สุดท้ายกำจัดค่าว่าง (NA) ที่เกิดจากการคำนวณวันแรก:

python



```
rets = log_return.dropna()
```

ผลลัพธ์ **rets** จะเป็น **DataFrame** ที่แสดง Log Return รายวันของหุ้นแต่ละตัว



Date	CPF.bk	CBG.bk	TU.bk	OSP.bk
01/03/2020	0.017858	-0.003016	0.021819	0.018349
01/06/2020	-0.017858	-0.021375	-0.051672	-0.006079
01/07/2020	0.008969	0.027399	0.022473	0.018127
01/08/2020	0.000000	-0.024318	-0.014926	0.011905
01/09/2020	0.052186	-0.003082	0.044125	0.040585

## ต่อจากนั้นคำนวณหาค่า Mean, Standard Deviation

จาก Log Return ของหุ้นแต่ละตัว:

- คำนวณค่า Mean และ Standard Deviation
  - คำนวณ **ค่าเฉลี่ย (Mean)** เพื่อตรวจสอบผลตอบแทนเฉลี่ยของหุ้นต่อวัน
  - คำนวณ **ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)** เพื่อประเมินความผันผวนของหุ้น
- คำนวณการแจกแจงปกติ (Normal Distribution)
  - ใช้ค่า Mean และ Std ที่คำนวณไว้มา สร้างฟังก์ชันความหนาแน่นของการแจกแจงปกติ (PDF) ซึ่งใช้เป็นการอ้างอิงสำหรับเปรียบเทียบการกระจายของผลตอบแทน
- วาดกราฟ Histogram และ Normal PDF

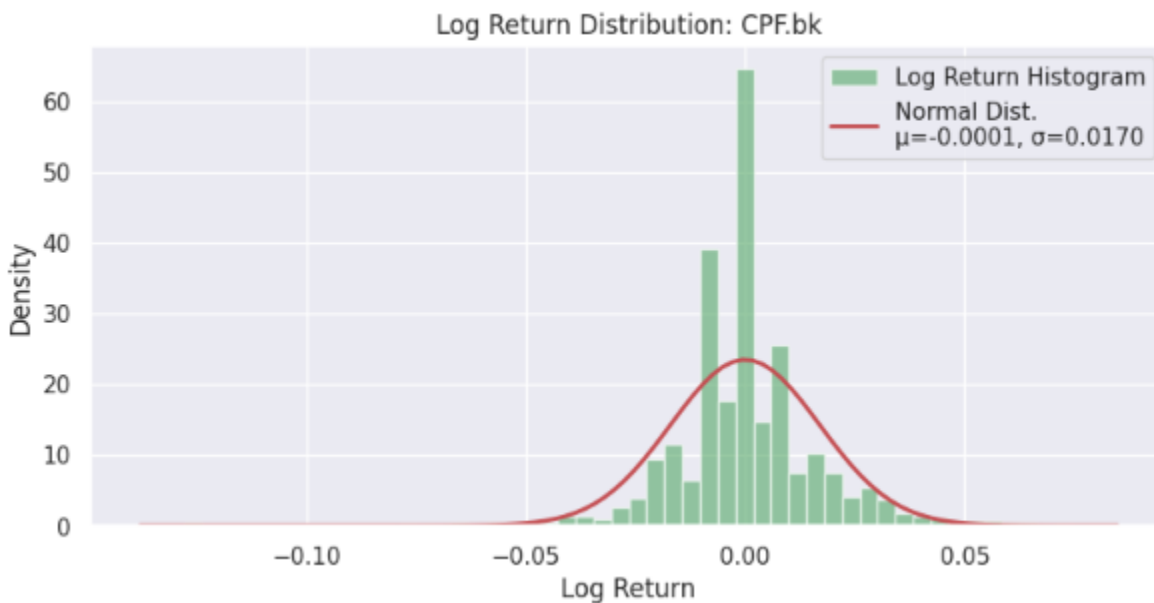
แสดงผลเป็นกราฟ:

แกน X: ผลตอบแทน Log Return

แกน Y: ความหนาแน่น (Density) ของความถี่

Histogram แสดงการกระจายจริงของ Log Return

เส้นโค้งสีแดง (Normal PDF) แสดงการกระจายปกติจากค่าเฉลี่ย  
และส่วนเบี่ยงเบนมาตรฐาน



### ข้อมูลพื้นฐานของหุ้น CPF.BK

- ค่าเฉลี่ยของ Log Return ( $\mu$ ) เท่ากับ -0.000059  
หมายความว่า โดยเฉลี่ยราคาหุ้น CPF จะลดลงประมาณ 0.0059% ต่อวัน  
(ถ้าค่าเป็นลบ แสดงว่าราคาหุ้นมีแนวโน้มลดลงเล็กน้อยในระยะยาว)
- ส่วนเบี่ยงเบนมาตรฐาน ( $\sigma$ ) เท่ากับ 0.017002  
แสดงถึงความผันผวนของราคาหุ้นในแต่ละวัน โดยราคาผันผวนประมาณ 1.7002% ต่อวัน  
(ค่ายิ่งสูง แปลว่าราคาหุ้นแกว่งตัวมากขึ้นและความเสี่ยงสูงขึ้น)

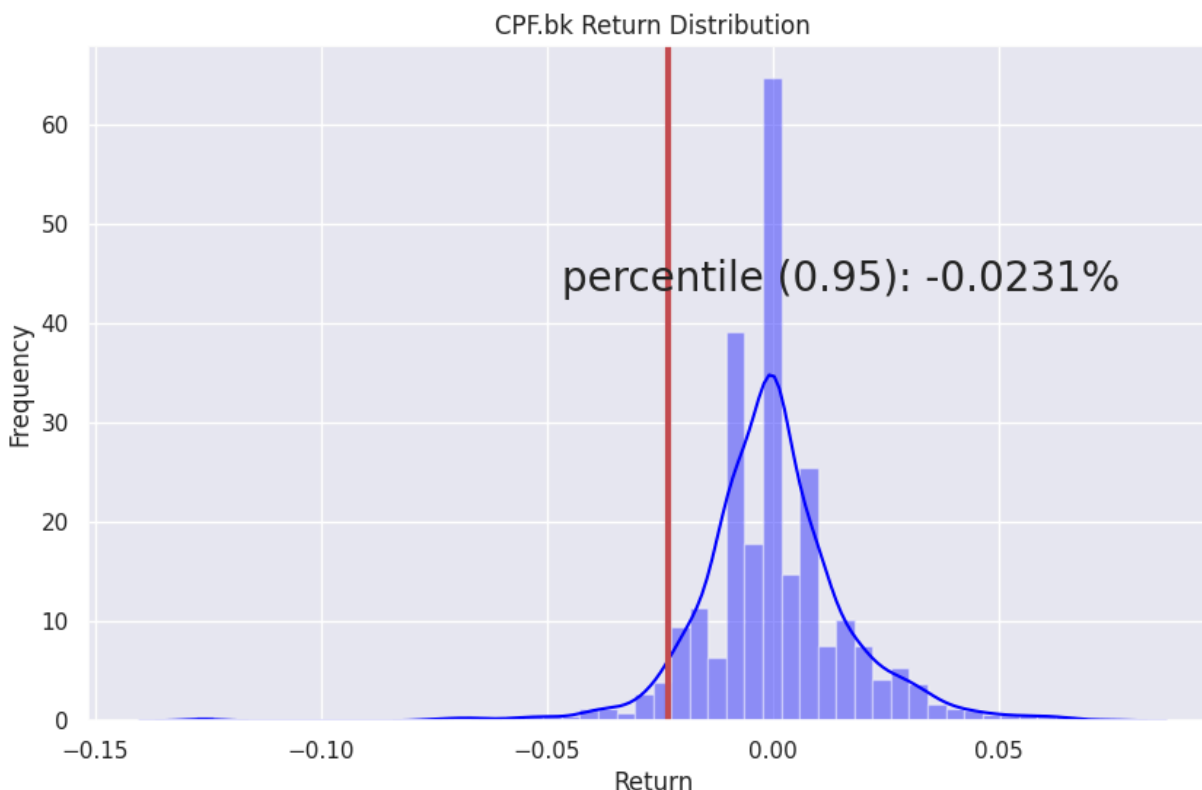
---

## สรุปการประเมินความเสี่ยงด้วย Percentile และ VaR -

### Down- Side

- **Percentile** คือค่าที่บ่งบอกตำแหน่งของข้อมูลในกลุ่มข้อมูลทั้งหมดในรูปแบบเปอร์เซ็นต์สะสม  
เช่น Percentile 5 หมายถึงค่าที่ต่ำกว่าข้อมูลทั้งหมด 5%  
หรือก็คือ ใน 100 วัน จะมีเพียง 5 วันที่ผลตอบแทนต่ำกว่าค่านี้ (คือวันที่แย่ที่สุด 5%)
- **Value at Risk (VaR)** ที่ระดับความมั่นใจ 95% หมายถึง  
การประเมินความเสียหายสูงสุดที่คาดว่าจะไม่เกินภายในช่วงเวลาหนึ่ง (เช่น 1 วัน)  
ซึ่งค่าดังกล่าวคือค่า Percentile 5 นั้นเอง  
หรือพูดง่าย ๆ คือ "มีโอกาสเพียง 5% ที่ผลตอบแทนจะลดลงต่ำกว่าค่านี้"

### ผลลัพธ์



1. ใช้ข้อมูลราคาหุ้น CPF ตั้งแต่วันที่ 1 มกราคม 2020 ถึง 31 ธันวาคม 2024  
รวมระยะเวลา 5 ปี

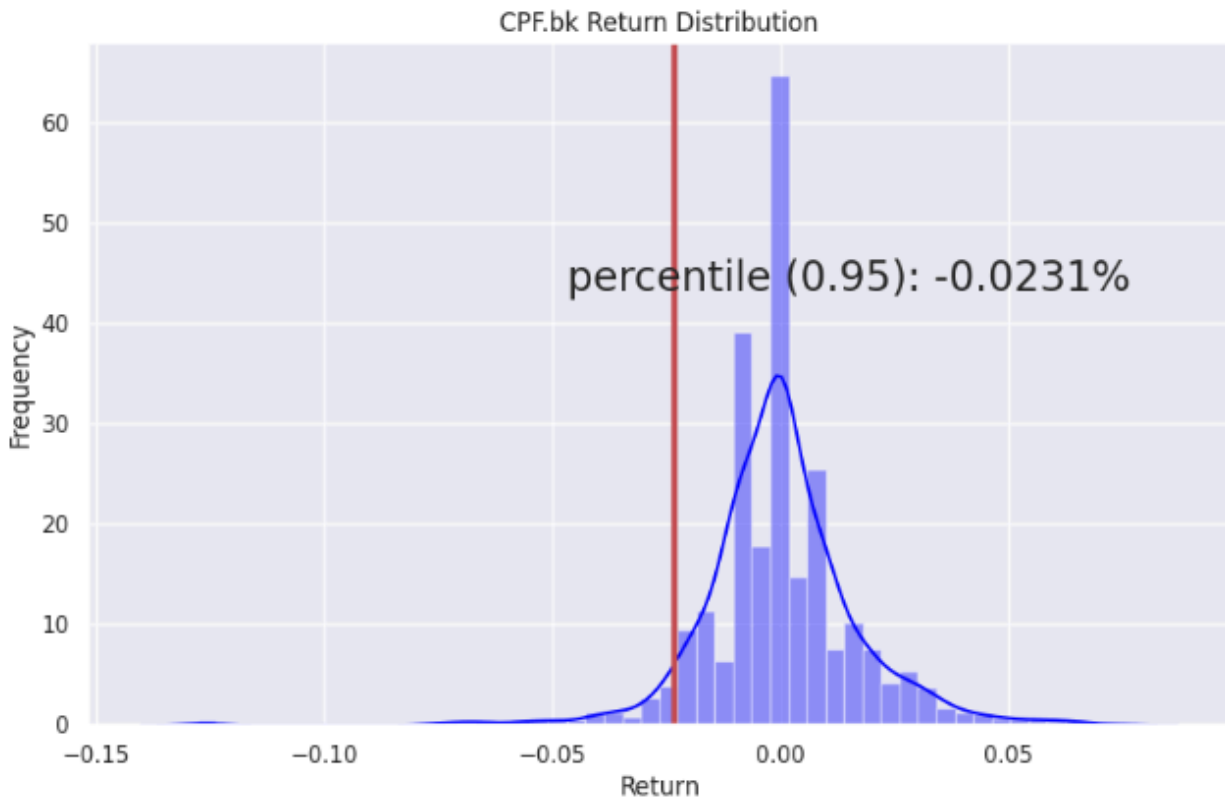
2. คำนวณค่า Percentile 5 ของผลตอบแทนรายวัน เพื่อดูว่าภายใน 5% ของวันที่แย่ที่สุด ราคาหุ้นจะลดลงเกินกว่าค่านี้หรือไม่

\*\*\*\* ผลลัพธ์พบว่า

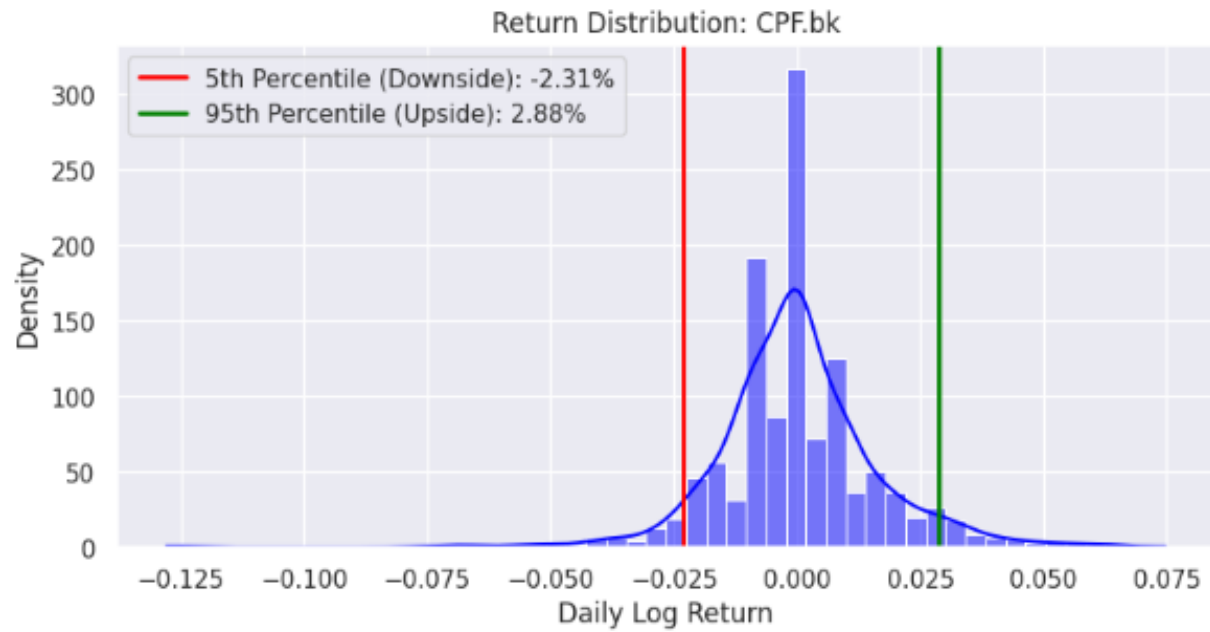
- ผลตอบแทนรายวันของหุ้น CPF ส่วนใหญ่อยู่ในช่วงที่ไม่ต่ำกว่า -2.31%
- มีเพียง 5% ของวันทั้งหมดในช่วง 5 ปีนี้ที่ราคาหุ้นลดลงเกิน -2.31%
- ค่านี้ช่วยบอกขอบเขตของความเสียด้านลบโดยอ้างอิงจากข้อมูลย้อนหลัง

- ทำให้เราประเมินความเสี่ยงของการขาดทุนในระดับนี้ได้อย่างชัดเจน

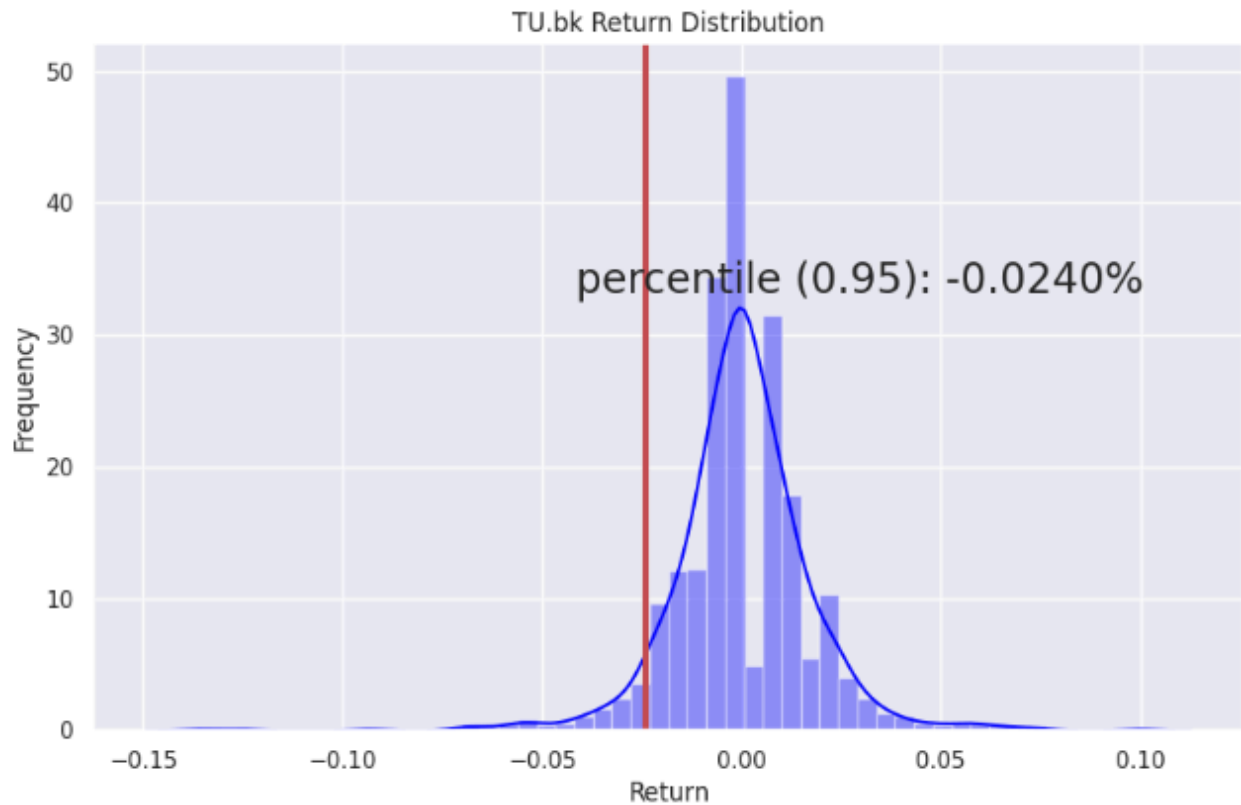
การเปรียบเทียบความเสี่ยงของหุ้นในกลุ่มเดียวกัน



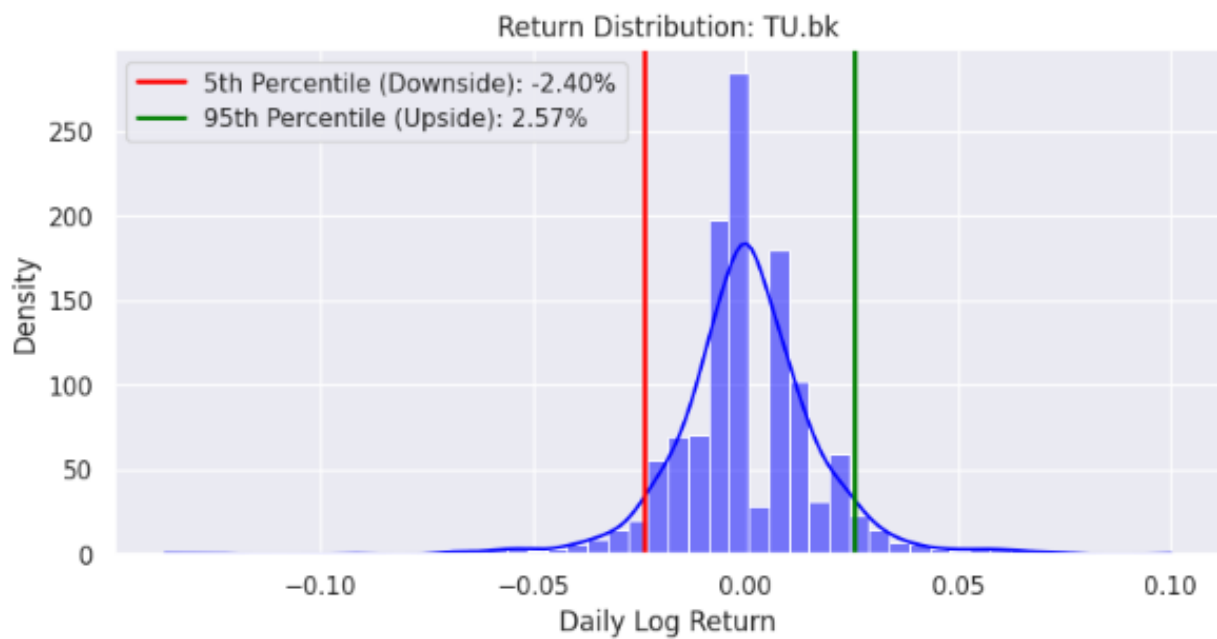
CPF: มีโอกาสเพียง 5% ที่ผลตอบแทนรายวันจะลดลงต่ำกว่า -2.31% ความเสี่ยงต่ำสุดในกลุ่ม



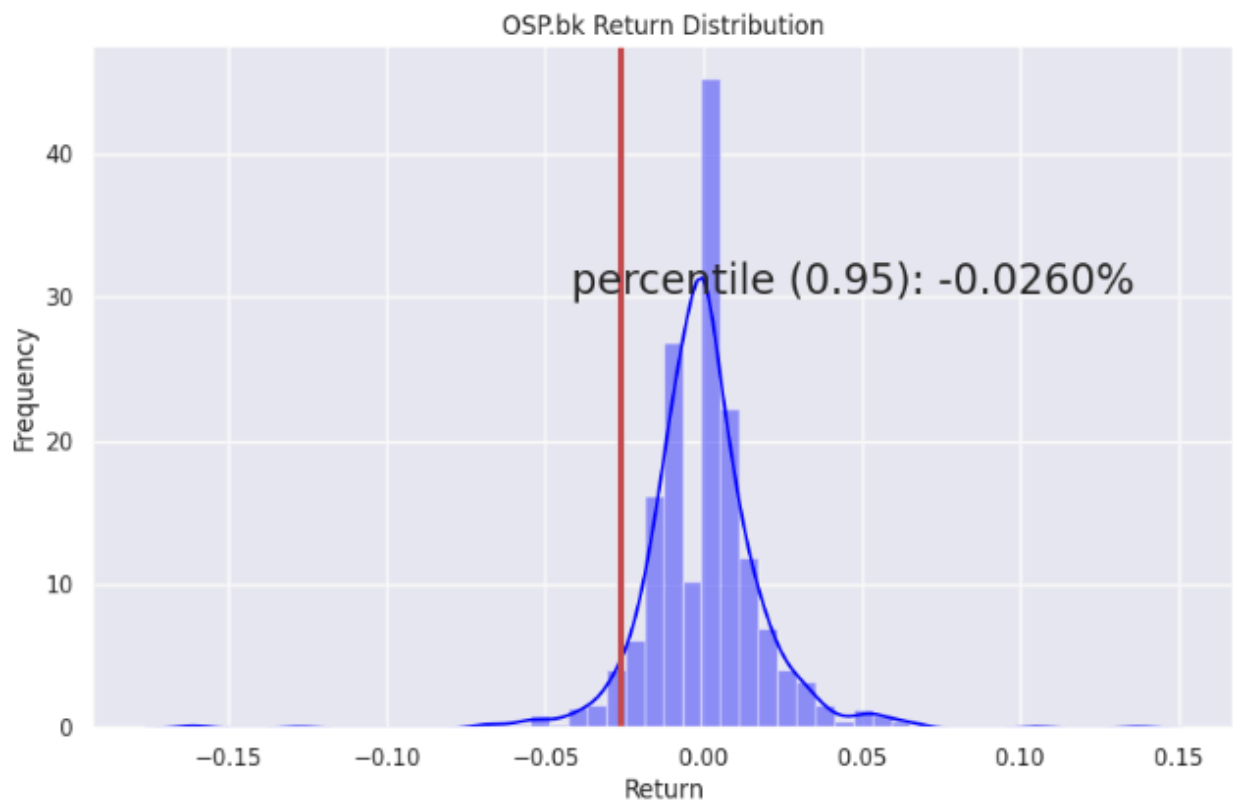
CPF: ความมั่นใจ 95%: มีโอกาส 5% ที่หุ้นจะขึ้นมากกว่า 2.88% ต่อวัน (Upside Potential)



TU: มีโอกาสเพียง 5% ที่ผลตอบแทนรายวันจะลดลงต่ำกว่า -2.40% ความเสี่ยงระดับกลาง

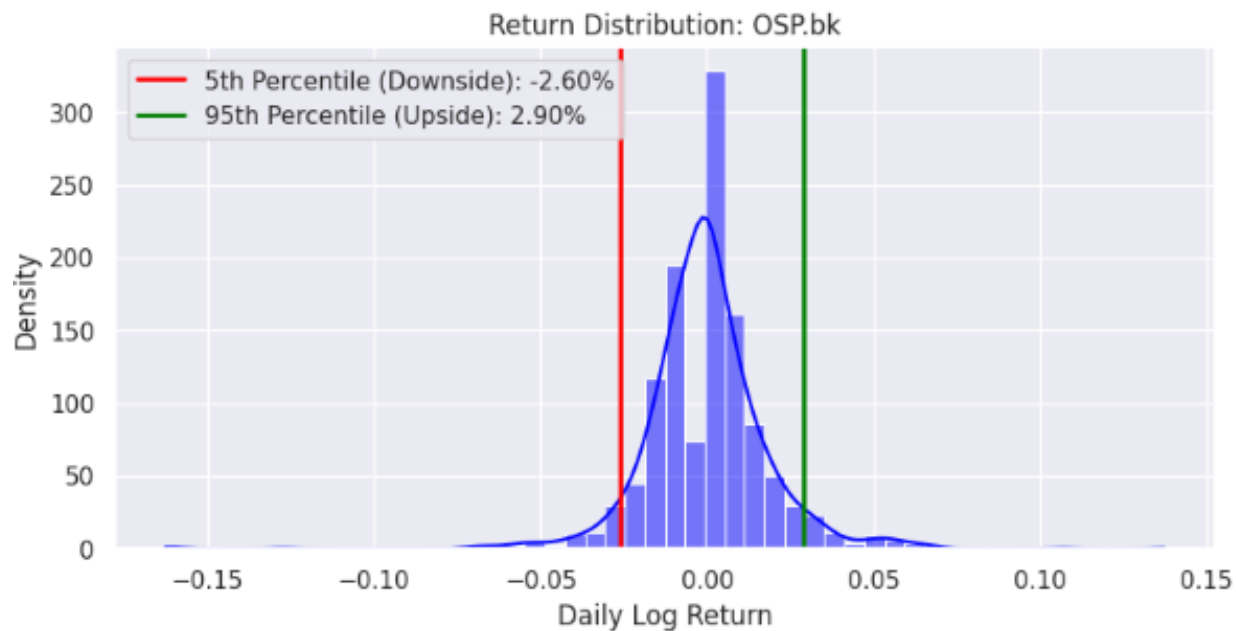


TU: ความมั่นใจ 95%: มีโอกาส 5% ที่หุ้นจะขึ้นมากกว่า 2.57% ต่อวัน (Upside Potential)

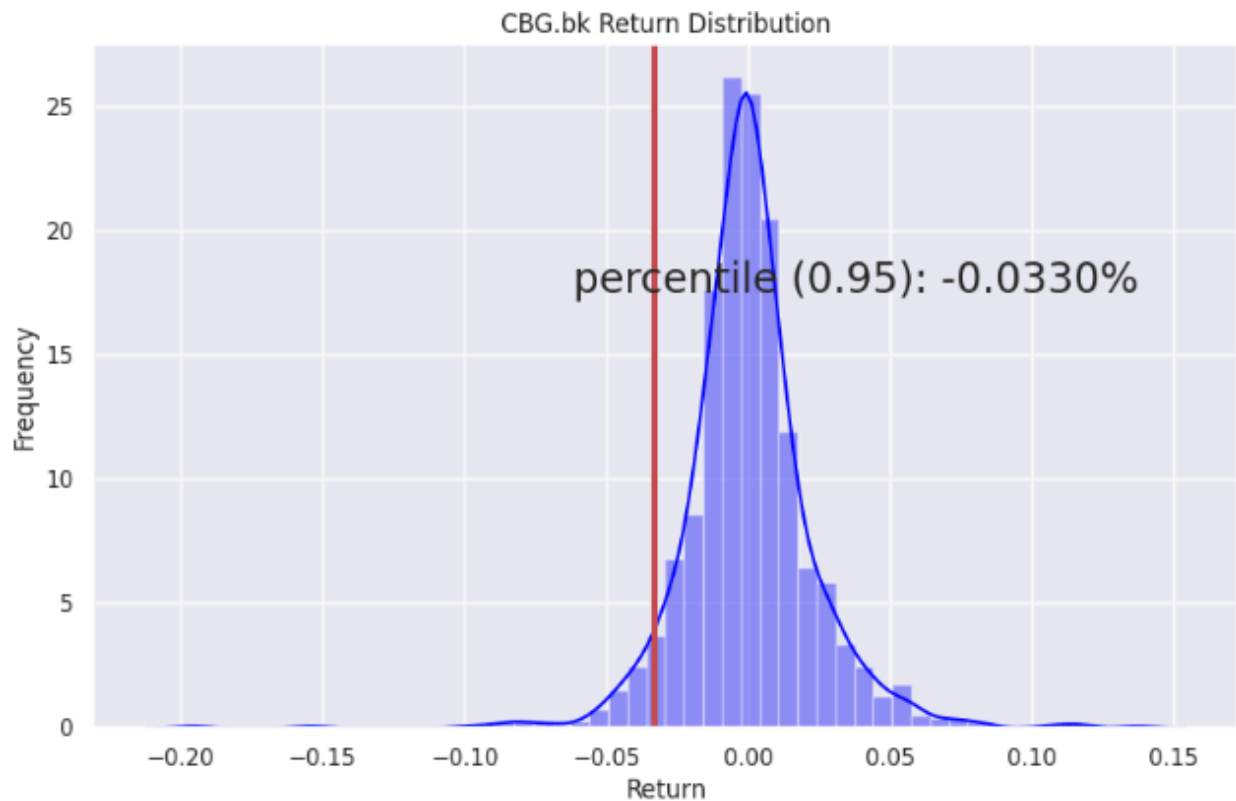


OSP: มีโอกาสเพียง 5% ที่ผลตอบแทนรายวันจะลดลงต่ำกว่า -2.60% ความเสี่ยงระดับกลาง-สูง

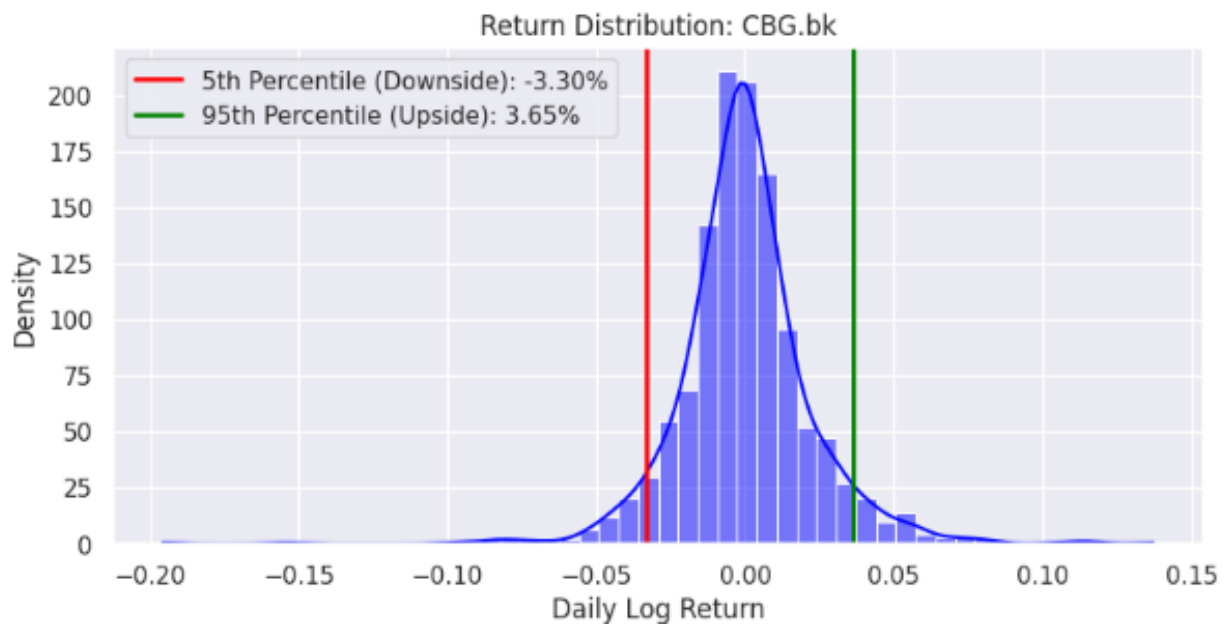




OSP: ความมั่นใจ 95%: มีโอกาส 5% ที่หุ้นจะขึ้นมากกว่า 2.90% ต่อวัน (Upside Potential)



CBG: มีโอกาสเพียง 5% ที่ผลตอบแทนรายวันจะลดลงต่ำกว่า -3.30% มีความเสี่ยงสูงสุดในกลุ่ม



CBG: ความมั่นใจ 95%: มีโอกาส 5% ที่หุ้นจะขึ้นมากกว่า 3.65% ต่อวัน  
(Upside Potential)