

การจัด portfolio ของหุ้นใน SET 50 ด้วย k-means และหาน้ำหนักที่เหมาะสมด้วย modern portfolio theory

บทนำ

ที่มาและความสำคัญ

การลงทุนในสินทรัพย์การเงินที่เป็นหุ้นนั้นเป็นหนึ่งในเครื่องมือหลักๆที่นักลงทุนมักจะเลือกลงทุน เนื่องจากมีผลตอบแทนเฉลี่ยที่มากกว่าการลงทุนอื่นๆ เช่น การลงทุนในอสังหาริมทรัพย์ การลงทุนในตราสารหนี้ และอาจจะมีผลตอบแทนเฉลี่ยที่น้อยกว่าตราสารอนุพันธ์ ทั้งนี้เป็นไปตามหลักการ ความเสี่ยงมาก ผลตอบแทนคาดหวังมาก ความเสี่ยงน้อย ผลตอบแทนคาดหวังก็น้อยตาม แต่ในทฤษฎีพอร์ตการลงทุนสมัยใหม่ (Modern Portfolio Theory) ของ Harry Markowitz ได้เสนอแนวคิดการกระจายความเสี่ยง โดยการถือครองสินทรัพย์เป็นกลุ่มที่มีความสัมพันธ์กันเองกันต่ำ (Low Correlation) ทำให้เมื่อเกิดสภาพเศรษฐกิจแบบหนึ่งในอนาคต หุ้นตัวหนึ่งอาจมีราคาที่สูง แต่หุ้นอีกตัวหนึ่งอาจมีราคาที่สูง ทำให้ผลตอบแทนโดยรวมไม่ผันผวนมากนัก ซึ่งเป็นการลดความเสี่ยงจากการลงทุนและทำให้พอร์ตโดยรวมได้ผลตอบแทนที่สูงขึ้น ณ ระดับความเสี่ยงหนึ่ง

โดยในการคัดเลือกหุ้นมาลงทุนนั้นเราจะใช้เทคนิค k-means clustering ในการจัดกลุ่มหุ้นให้เราแทนที่จะใช้วิธีการคัดเลือกหุ้นที่อยู่คนละอุตสาหกรรมมาอยู่ในพอร์ตและใช้การวิเคราะห์ของนักวิเคราะห์การเงินในการคัดเลือกหุ้นมาจัดพอร์ต ซึ่งอาจมีความผิดพลาดจากการวิเคราะห์ที่ได้ โดยเมื่อเราคัดเลือกหุ้นเป็นกลุ่มได้แล้ว เราจะได้ว่าหุ้นที่อยู่ในกลุ่มเดียวกันจะมีความคล้ายหรือมีความสัมพันธ์กันเองเยอะ (High Correlation) และหุ้นที่อยู่ต่างกลุ่มกันจะมีความคล้ายหรือมีความสัมพันธ์กันเองต่ำ (Low Correlation) ทำให้เราตัดสินใจเลือกหุ้นมาแค่ 1 หุ้นในแต่ละกลุ่มเพื่อป้องกันการมีความสัมพันธ์ระหว่างกันที่สูงของหุ้นที่อยู่ในพอร์ตเดียวกัน โดยเกณฑ์ที่ใช้ในการเลือกหุ้นในแต่ละกลุ่ม คือ ค่า Sharpe ratio ซึ่งเป็นค่าที่บอกถึงผลตอบแทน ณ ระดับความเสี่ยงหนึ่ง แปลว่าหุ้นที่มีค่า Sharpe ratio สูงกว่าหุ้นอื่นๆ จะมีความผลตอบแทนที่สูงกว่า ณ ระดับความเสี่ยงที่เท่ากัน ดังนั้นเราจึงใช้เกณฑ์นี้ในการคัดเลือกหุ้นที่มีค่า Sharpe ratio สูงที่สุดในแต่ละกลุ่มมาจัดพอร์ต โดยการสุ่มน้ำหนักของหุ้นในพอร์ต 100,000 ครั้งเพื่อเอามาพล็อตกราฟเส้น efficient frontier หาน้ำหนักที่เหมาะสมที่สุด 2 พอร์ต คือ น้ำหนักที่ทำให้พอร์ตมีผลตอบแทนระดับหนึ่งที่ความเสี่ยงต่ำสุด กับ น้ำหนักที่ทำให้พอร์ตมีผลตอบแทนสูงสุดที่ความเสี่ยงระดับหนึ่ง จากนั้นนำทั้ง 2 พอร์ตมาหาผลตอบแทนสะสม และหาผลตอบแทนสะสมของพอร์ตที่มีน้ำหนักที่เท่ากันด้วย แล้วเอามาพล็อตกราฟเปรียบเทียบกับผลตอบแทนสะสมของ SET index เพื่อดูลักษณะของกราฟที่มีน้ำหนักทั้ง 3 แบบ

จุดประสงค์ของงานวิจัย

- เพื่อนำหลักการของทฤษฎีพอร์ตการลงทุนสมัยใหม่ (Modern Portfolio Theory) และ k-means clustering มาประยุกต์ใช้ในการจัดพอร์ต
- เพื่อศึกษาเปรียบเทียบลักษณะของผลตอบแทนสะสมของพอร์ตที่มีน้ำหนักการลงทุนตามความเสี่ยงต่ำที่สุด ผลตอบแทนสะสมของพอร์ตที่มีน้ำหนักการลงทุนตามผลตอบแทนสูงสุด และผลตอบแทนสะสมของพอร์ตที่มีน้ำหนักการลงทุนเท่ากัน

ประโยชน์ของงานวิจัย

- เป็นแนวทางใหม่ในการจัดพอร์ตหุ้นที่มีการนำหลักการทฤษฎีพอร์ตการลงทุนสมัยใหม่ (Modern Portfolio Theory) k-means clustering การใช้ Sharpe ratio ในการคัดเลือกหุ้นในแต่ละกลุ่ม รวมถึงการหาน้ำหนักที่เหมาะสมของพอร์ตด้วยการสุ่มมาใช้ในการจัดพอร์ต

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

k-means clustering algorithm

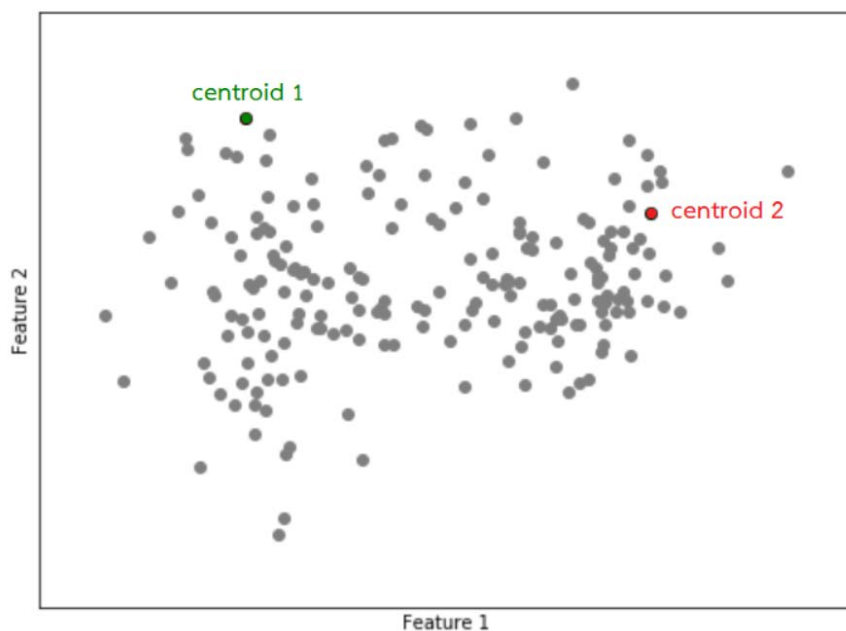
เป็นเทคนิคใน Data Science ที่จะมาใช้ในการแก้ปัญหาต่าง ๆ โดยอาศัยข้อมูลนั้นมีหลากหลายวิธี และในบทความนี้จะพูดถึงวิธีการใช้เทคนิค machine learning ในการแบ่งกลุ่มข้อมูล หรือ clustering algorithm ซึ่งเชื่อว่าโมเดลหลาย ๆ คนน่าจะนึกถึงเป็นอันดับแรกก็คือ k-means เพราะด้วยความเรียบง่ายของโมเดล และประสิทธิภาพของการแบ่งกลุ่มที่อยู่ในเกณฑ์ที่ยอมรับได้ k-means จึงเป็นที่นิยมและมักจะถูกเรียกใช้งานอยู่บ่อยครั้ง

โดยทั่วไปแล้ว k-means มีวิธีการทำงานโดยเริ่มจากการกำหนดจุดศูนย์กลางของกลุ่ม (centroid) มาจำนวน k จุดโดยที่ k คือจำนวนกลุ่มที่คาดว่าจะได้จากการแบ่งกลุ่ม จากนั้นคำนวณระยะห่างระหว่างข้อมูลในแต่ละแถวกับ centroid เพื่อใช้ในการจัดกลุ่มข้อมูล โดยระยะห่างนั้นสามารถคำนวณโดยใช้ Euclidean distance

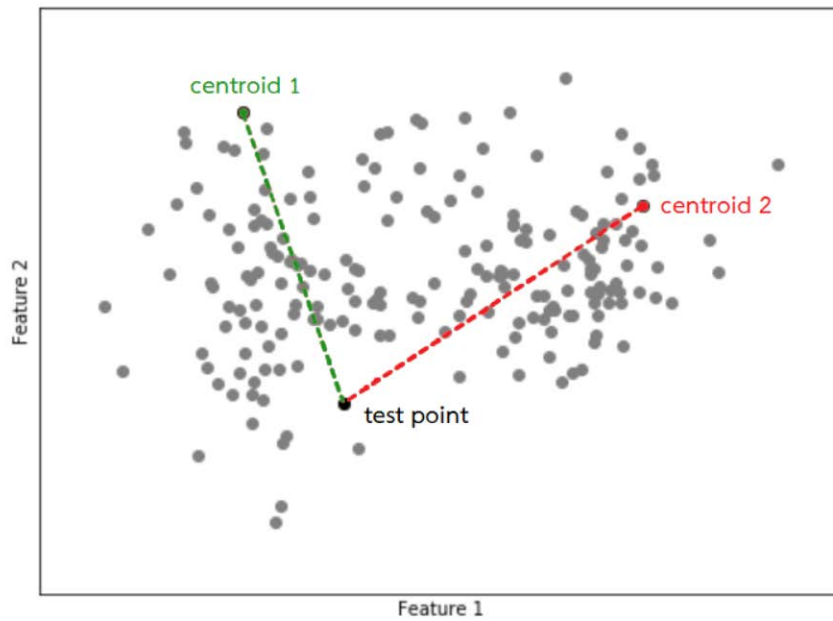
$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

โดยที่ x และ y เป็นข้อมูลแต่ละแถวที่มีความยาว n หรือมี n ค่า ซึ่งสามารถเข้าถึงแต่ละค่าได้จาก index i เมื่อจัดกลุ่มครบทุกแถวข้อมูลแล้วจะมีการปรับตำแหน่งของ centroid ในแต่ละกลุ่มใหม่ และจะทำวนซ้ำไปเรื่อย ๆ จนกว่าจะตรงตามเงื่อนไขในการหยุดการทำงาน ซึ่งสามารถสรุปเป็นขั้นตอนได้ดังนี้

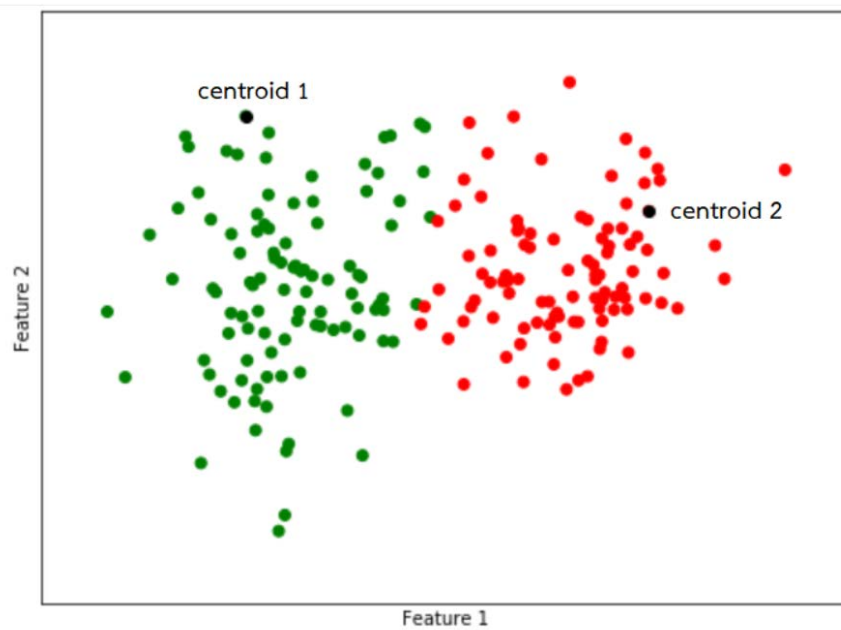
1. สุ่มข้อมูลมา k records เพื่อใช้เป็น centroid (ในกรณีนี้ $k=2$)



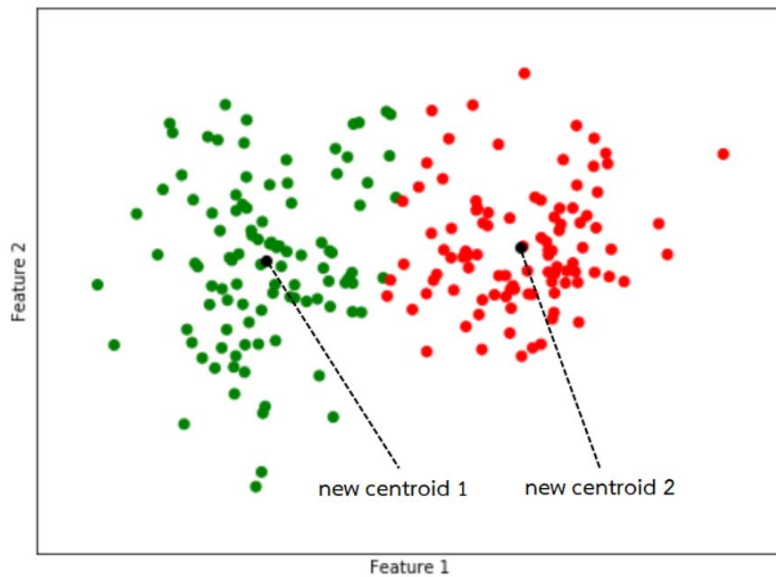
2. คำนวณระยะห่างระหว่างข้อมูลแต่ละแถวเทียบกับ centroid ที่มีอยู่ k จุด



3. จัดกลุ่มจุดข้อมูลให้ไปอยู่กลุ่มเดียวกับ centroid ที่อยู่ใกล้ที่สุด



4. ปรับตำแหน่งของ centroid ในแต่ละกลุ่มไปเป็นค่าเฉลี่ยของข้อมูลทั้งหมดที่อยู่ในกลุ่มนั้น



5. ทำขั้นตอนที่ 2 ถึง 4 จนกว่าจะถึงเกณฑ์ที่ใช้ในการหยุด

แม้ว่าจะเป็นโมเดลที่ได้รับความนิยมมากหรือมีความเรียบง่ายแค่ไหนก็ตาม k-means มีข้อจำกัดอย่างหนึ่งคือสามารถใช้ในการแบ่งกลุ่มข้อมูลที่เป็นตัวเลข (numerical feature) ได้เท่านั้น เลยทำให้เกิดคำถามที่ว่า “จะต้องทำอย่างไรหากต้องการแบ่งกลุ่มข้อมูลที่ไม่ใช่ตัวเลข หรือข้อมูลที่เป็นหมวดหมู่ (categorical feature)” เนื่องจากว่าในวิธีการทำงานของ k-means มีการคำนวณระยะห่าง ซึ่งถ้าข้อมูลไม่ได้มีแค่ตัวเลข แต่มีตัวอักษรหรือข้อความรวมอยู่ด้วย จะไม่สามารถทำในขั้นตอนที่ 2 และ 4 ได้ เพราะระยะห่างและค่าเฉลี่ยระหว่างตัวอักษรหรือข้อความนั้นไม่สามารถถูกคำนวณได้ด้วยสูตรแบบเดียวกันกับที่ใช้ในการคำนวณข้อมูลที่เป็นตัวเลข อย่างไรก็ตาม Z. Huang¹ ได้เอาชนะข้อจำกัดนี้ โดยเปลี่ยนการวัดระยะห่างสำหรับ categorical feature ใหม่ เป็นการวัดความคล้ายคลึงระหว่างกันแทน ซึ่งกำหนดว่าหากเป็นข้อมูลที่มีค่าเดียวกันจะให้ค่าเป็น 0 และถ้าเป็นค่าที่แตกต่างกันจะให้ค่าเป็น 1 และเรียกปริมาณนี้ว่า dissimilarity measure

$$\delta(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{x} \neq \mathbf{y} \\ 0 & \text{if } \mathbf{x} = \mathbf{y} \end{cases}$$

โดยที่ \mathbf{x} และ \mathbf{y} เป็นตัวอักษรหรือข้อความใด ๆ ด้วยเหตุนี้จึงทำให้ปริมาณที่ใช้วัด “ระยะห่าง” ระหว่างข้อมูลแต่ละตัวกับจุด centroid ในแต่ละกลุ่มจะถูกเปลี่ยนแปลงตามไปด้วย โดยจะกลายเป็นส่วนผสมระหว่างระยะห่างจริง ๆ ที่เคยใช้ใน k-means หรือก็คือ Euclidean distance เพื่อคำนวณสำหรับ numerical feature และระยะหรือความคล้ายคลึงที่สร้างขึ้นใหม่ หรือก็คือ dissimilarity measure เพื่อคำนวณสำหรับ categorical feature ซึ่งสามารถเขียนเป็นสมการได้เป็นดังนี้

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i \in n_r} (x_i - y_i)^2 + \gamma \sum_{i \in n_c} \delta(x_i, y_i)$$

โดยที่ x และ y เป็นข้อมูลแต่ละแถว โดยมี numerical feature อยู่ n_r คอลัมน์ และมี categorical feature อยู่ n_c คอลัมน์ และ δ คือ weight ของ dissimilarity measure

เมื่อทำการแทนที่ระยะห่างใน k-means ด้วยปริมาณดังกล่าวก็ทำให้วิธีการในขั้นตอนที่ 2 กลับมาทำงานได้อีกครั้ง สำหรับขั้นตอนที่ 4 ที่มีการปรับตำแหน่งของ centroid ในแต่ละกลุ่มจะถูกแยกคิด ถ้าเป็น numerical feature ยังคงใช้ค่าเฉลี่ยของ record เหมือนเดิม แต่ถ้าเป็น categorical feature จะเปลี่ยนไปใช้ค่าฐานนิยม (mode) ของ record ทั้งหมดที่อยู่ในกลุ่มนั้นแทน โดยวิธีการที่ถูกพัฒนาขึ้นมาใหม่นี้มีชื่อเรียกว่า k-prototypes

แท้จริงแล้ว k-prototypes คืออีกรูปแบบหนึ่งของ k-means ที่ถูกพัฒนาให้รองรับ categorical feature ซึ่งถ้าข้อมูลที่ใช้เป็น numerical feature ทั้งหมด k-prototypes จะถูกเรียกว่า k-means และถ้าข้อมูลที่ใช้เป็น categorical feature ทั้งหมด k-prototypes จะถูกเรียกว่า k-modes

ทฤษฎีกลุ่มหลักทรัพย์ของ Markowitz

แนวคิดตามทฤษฎีของ Markowitz เป็นแนวคิดที่เริ่มว่า การกระจายการลงทุนช่วยลดความเสี่ยงเฉพาะในกรณีที่เป็นการลงทุนเป็นกลุ่มหลักทรัพย์ ที่หลักทรัพย์แต่ละคู่ไม่ได้มีความสัมพันธ์ในลักษณะที่ไปด้วยกันอย่างสมบูรณ์ (ค่าสัมประสิทธิ์สหสัมพันธ์ต่ำกว่า +1.0) จึงสามารถลดค่าเบี่ยงเบนมาตรฐานของกลุ่มหลักทรัพย์ลงได้ แต่ถ้ากระจายการลงทุนในหลักทรัพย์หลายชนิดที่มีลักษณะความสัมพันธ์ระหว่างอัตราผลตอบแทนที่ไปด้วยกันอย่างสมบูรณ์ (ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ +1.0) จะไม่สามารถลดความเสี่ยงของกลุ่มหลักทรัพย์ลงได้

นอกจากนี้ ทฤษฎีกลุ่มหลักทรัพย์ของ Markowitz ได้แสดงให้เห็นว่า ผู้ลงทุนสามารถสร้างกลุ่มหลักทรัพย์ต่างๆ ที่ให้อัตราผลตอบแทนที่คาดหวัง และค่าส่วนเบี่ยงเบนมาตรฐานของอัตราผลตอบแทนของกลุ่มหลักทรัพย์ในระดับต่างๆ ได้ ทั้งนี้จะมีกลุ่มหลักทรัพย์ต่างๆ จำนวนหนึ่งที่เหนือกว่าหรือมีประสิทธิภาพกว่ากลุ่มหลักทรัพย์อื่นๆ กล่าวคือ เมื่อพิจารณา ณ ความเสี่ยงระดับหนึ่ง กลุ่มหลักทรัพย์เหล่านี้เป็นกลุ่มหลักทรัพย์ที่ให้อัตราผลตอบแทนสูงสุด ในทำนองเดียวกัน ณ อัตราผลตอบแทนระดับหนึ่ง กลุ่มหลักทรัพย์เหล่านี้เป็นกลุ่มหลักทรัพย์ที่มีความเสี่ยงที่ต่ำที่สุด ตามทฤษฎีกลุ่มหลักทรัพย์ของ Markowitz เรียกแนวขอบที่กลุ่มหลักทรัพย์นี้เรียงตัวกันอยู่ว่า “เส้นโค้งกลุ่ม 13 หลักทรัพย์ที่มีประสิทธิภาพ” (Efficient Frontier) ผู้ลงทุนจะเลือกลงทุนในกลุ่มหลักทรัพย์ที่มีประสิทธิภาพตามระดับความเสี่ยงและอัตราผลตอบแทนที่ผู้ลงทุนต้องการ

ข้อสมมติฐานตามแนวความคิดการสร้างกลุ่มหลักทรัพย์ของ Markowitz อยู่ภายใต้ ข้อสมมติฐานอันเกี่ยวกับพฤติกรรมของผู้ลงทุนดังต่อไปนี้

1. การตัดสินใจลงทุนในแต่ละทางเลือก ผู้ลงทุนจะพิจารณาจากการกระจายของโอกาสที่จะเกิดผลตอบแทนของกลุ่มหลักทรัพย์ในช่วงระยะเวลาลงทุน
2. ผู้ลงทุนจะพยายามทำให้อรรถประโยชน์ที่คาดว่าจะได้รับต่อ 1 งวดเวลาลงทุนสูงที่สุดโดยเส้นอรรถประโยชน์ของผู้ลงทุนแสดงถึงอรรถประโยชน์ที่เพิ่มขึ้นในอัตราที่ลดลง เมื่อมีความมั่งคั่งสูงขึ้น
3. ผู้ลงทุนแต่ละรายจะกำหนดความเสี่ยงจากการลงทุนบนพื้นฐานของความแปรปรวนของอัตราผลตอบแทนที่คาดว่าจะได้รับ

4. การตัดสินใจของผู้ลงทุนขึ้นกับอัตราผลตอบแทนที่คาดว่าจะได้รับและความเสี่ยงเท่านั้น ดังนั้น เส้นอรรถประโยชน์จึงเป็นฟังก์ชันของอัตราผลตอบแทนที่คาดหวัง กับค่าที่คาดหวังของความแปรปรวนหรือส่วนเบี่ยงเบนมาตรฐานของอัตราผลตอบแทน

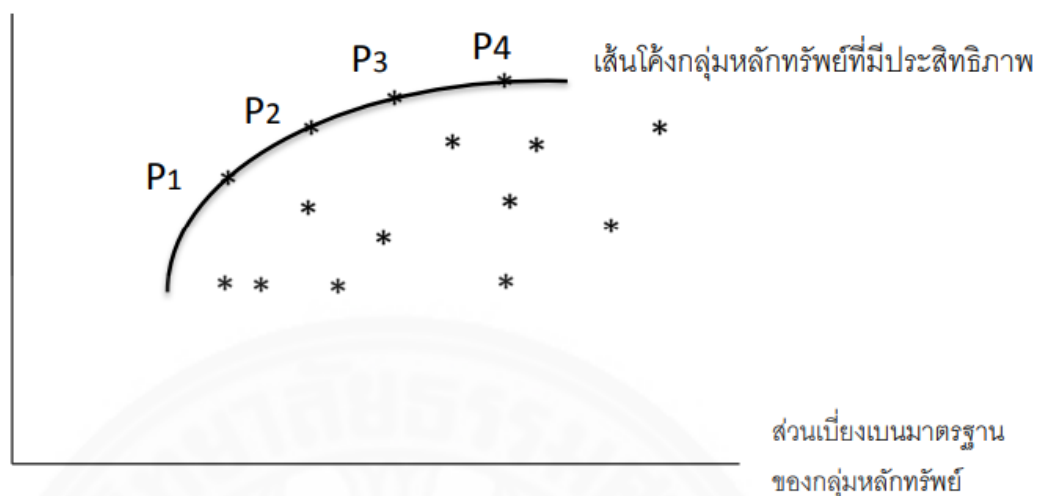
5. ภายใต้ระดับความเสี่ยงหนึ่ง ผู้ลงทุนจะเลือกการลงทุนที่ให้อัตราผลตอบแทนสูงสุด ในทำนองเดียวกัน ภายใต้อัตราผลตอบแทนระดับหนึ่ง ผู้ลงทุนจะเลือกการลงทุนที่มีความเสี่ยงต่ำสุด

เส้นโค้งกลุ่มหลักทรัพย์ที่มีประสิทธิภาพ (Efficient Frontier)

จากกลุ่มหลักทรัพย์ที่ประกอบด้วยหลักทรัพย์เพียง 2 ชนิด เมื่อสมมติให้หลักทรัพย์ทั้งสองมี ค่าสัมประสิทธิ์สหสัมพันธ์ต่างๆ จะสามารถสร้างกลุ่มหลักทรัพย์ได้มากมายตามสัดส่วนของเงินลงทุนที่เปลี่ยนไป ดังนั้น หากผู้ลงทุนคัดเลือกหลักทรัพย์ที่สอดคล้องกับนโยบายการลงทุนได้หลักทรัพย์มาจำนวนหนึ่ง ผู้ลงทุนสามารถสร้างกลุ่มหลักทรัพย์ที่มีค่าอัตราผลตอบแทนและความเสี่ยงที่หลากหลายตามตัวแปรต่างๆ คือ

1. จำนวนหลักทรัพย์ที่ประกอบขึ้นเป็นกลุ่มหลักทรัพย์
2. ค่าเบี่ยงเบนมาตรฐานของอัตราผลตอบแทนของหลักทรัพย์ในกลุ่มหลักทรัพย์
3. ความสัมพันธ์ระหว่างความผันผวนของอัตราผลตอบแทนของหลักทรัพย์ในกลุ่มหลักทรัพย์
4. สัดส่วนของเงินลงทุนในแต่ละหลักทรัพย์

ผลตอบแทนที่คาดหวังของกลุ่มหลักทรัพย์



จากภาพแสดงอัตราผลตอบแทนที่คาดหวังและค่าเบี่ยงเบนมาตรฐานของกลุ่มหลักทรัพย์ต่างๆ ที่เป็นไปได้ ที่สามารถสร้างขึ้นได้ตามความหลากหลายของจำนวนหลักทรัพย์ ค่าส่วนเบี่ยงเบนมาตรฐานของหลักทรัพย์ ค่าสัมประสิทธิ์สหสัมพันธ์ และสัดส่วนเงินลงทุนในหลักทรัพย์แต่ละชนิดที่ประกอบกันเป็นกลุ่มหลักทรัพย์ กลุ่มหลักทรัพย์ P1, P2, P3 และ P4 เป็นกลุ่มหลักทรัพย์ที่ให้อัตราผลตอบแทนสูงสุด ณ ระดับความเสี่ยงหนึ่ง หรือให้ความเสี่ยงต่ำสุด ณ ระดับอัตราผลตอบแทนหนึ่ง กลุ่มหลักทรัพย์เหล่านี้ เรียกว่า “กลุ่มหลักทรัพย์ที่มีประสิทธิภาพ” (Efficient Portfolios) ผู้ลงทุนสามารถจัดสรรเงินลงทุนระหว่างกลุ่มหลักทรัพย์ที่มีประสิทธิภาพและสามารถสร้างกลุ่มหลักทรัพย์ที่มีประสิทธิภาพอีกจำนวนมากจนอาจกลายเป็นเส้นเชื่อมจุดแสดงอัตราผลตอบแทนและความเสี่ยงของกลุ่มหลักทรัพย์ที่มีประสิทธิภาพได้ และจะเรียกเส้นนี้ว่า “เส้นโค้งกลุ่มหลักทรัพย์ที่มีประสิทธิภาพ” (Efficient Frontier)

การวัดผลการดำเนินงานของพอร์ตการลงทุนด้วย Sharpe Index

William Sharpe ได้นำเสนอการคำนวณการเปรียบเทียบอัตราผลตอบแทนและความเสี่ยงของหลักทรัพย์ตัวใดตัวหนึ่งกับค่าดัชนีที่นิยมใช้ในตลาดนั้นๆ เรียกว่า Single – Index Model ซึ่งแนวคิดของ Sharpe มีสมมติฐานว่าความผันผวนของอัตราผลตอบแทนของหลักทรัพย์แต่ละชนิด ไม่ได้ขึ้นอยู่กับลักษณะเฉพาะของหลักทรัพย์นั้นเพียงอย่างเดียว แต่ยังขึ้นอยู่กับความผันผวนของอัตราผลตอบแทนโดยทั่วไปของตลาดที่ได้รับผลจากการเปลี่ยนแปลงของภาวะเศรษฐกิจและภาวะธุรกิจ โดยการเปลี่ยนแปลงของอัตราผลตอบแทนส่งผลต่อความพึงพอใจในระดับราคาของหลักทรัพย์และการเปลี่ยนแปลงของราคาหลักทรัพย์จะถูกแสดงออกในลักษณะของดัชนีราคา ดังนั้น Sharpe จึงคำนวณจากผลต่างระหว่างอัตราผลตอบแทนเฉลี่ยของกลุ่มสินทรัพย์กับอัตราผลตอบแทนที่ไม่มีความเสี่ยง (Rf)หารด้วยความเสี่ยงรวมของผลตอบแทนจากกลุ่มหลักทรัพย์ลงทุน ดังสมการ

$$\text{sharpe ratio} = \frac{Rp - Rf}{\sigma_i}$$

R_p = อัตราผลตอบแทนเฉลี่ยของพอร์ตการลงทุน

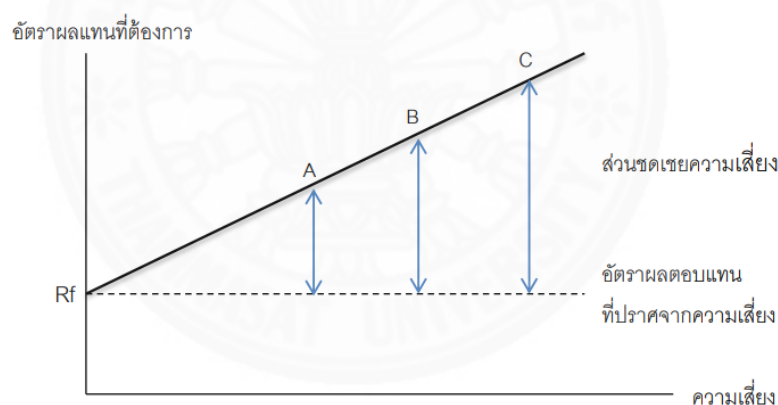
R_f = อัตราผลตอบแทนจากการลงทุนในหลักทรัพย์ที่ไม่มีความเสี่ยง

σ_i = ความเสี่ยงของพอร์ตการลงทุน

Sharpe Index เป็นดัชนีที่ใช้การปรับมาตรฐานของอัตราผลตอบแทนต่อความเสี่ยงทั้งหมดที่เกิดขึ้นให้อยู่บนพื้นฐานเดียวกัน (Risk-adjusted basis) จึงสามารถใช้เปรียบเทียบในการบริหารสินทรัพย์ของพอร์ตการลงทุนที่มีอัตราผลตอบแทนและความเสี่ยงที่แตกต่างกันได้ หากพอร์ตการลงทุนมีค่า Sharpe Index สูง แสดงว่ามีความสามารถในการบริหารหลักทรัพย์สูงกว่ากลุ่มสินทรัพย์หรือกองทุนที่มีค่า Sharpe Index ต่ำ เพราะมีอัตราผลตอบแทนต่อหนึ่งหน่วยความเสี่ยงที่สูงกว่า

ความสัมพันธ์ระหว่างอัตราผลตอบแทนที่ต้องการและความเสี่ยง

ตั้งอยู่บนสมมติฐานที่ว่าผู้ลงทุนแต่ละรายเป็นผู้ที่ไม่ชอบความเสี่ยงหรือหลีกเลี่ยงความเสี่ยง (Risk Averse) หากมีการลงทุนที่มีความเสี่ยงมากขึ้นผู้ลงทุนย่อมต้องการส่วนชดเชยความเสี่ยงที่มากขึ้น ทำให้ระดับอัตราผลตอบแทนที่ต้องการสูงขึ้น อัตราผลตอบแทนและความเสี่ยงจึงมีความสัมพันธ์ไปในทิศทางเดียวกัน



ภาพที่ 2.1 ความสัมพันธ์ระหว่างอัตราผลตอบแทนที่ต้องการกับความเสี่ยง

จากภาพที่ 2.1 แกนตั้งคืออัตราผลตอบแทนที่ต้องการของผู้ลงทุน และแกนนอนคือระดับความเสี่ยงของหลักทรัพย์ ยิ่งหลักทรัพย์มีความเสี่ยงสูง ผู้ลงทุนจะต้องการส่วนชดเชยความเสี่ยงมากขึ้น ทำให้อัตราผลตอบแทนที่ต้องการสูงขึ้น ดังนั้น หลักทรัพย์ที่มีความเสี่ยงสูงควรให้อัตราผลตอบแทนที่คาดหวังสูงเพื่อให้

ผู้ลงทุนพอใจ โดยที่ผู้ลงทุนแต่ละคนมีความพอใจในระดับอัตรา Rf A B C 7 ผลตอบแทนและความเสี่ยงที่แตกต่างกันไป นั่นคือ หากเป็นผู้ลงทุนประเภทที่ไม่ชอบความเสี่ยง ผู้ลงทุนจะเลือกลงทุนในหลักทรัพย์ที่มีความเสี่ยงต่ำในการจัดกลุ่มหลักทรัพย์เพื่อการลงทุน (Portfolio) ดังนั้น อัตราผลตอบแทนที่ผู้ลงทุนต้องการจะอยู่ที่จุด A แต่หากเป็นผู้ลงทุนที่ชอบความเสี่ยง หลักทรัพย์ที่ผู้ลงทุนจะเลือกเพื่อสร้างกลุ่มสินทรัพย์ลงทุน ก็จะเป็นหลักทรัพย์ที่มีความเสี่ยงสูง ส่งผลให้อัตราผลตอบแทนที่ผู้ลงทุนต้องการจะเลื่อนไปอยู่ ณ จุด C นอกจากนี้ หากผู้ลงทุนมีการเลือกหลักทรัพย์ที่จะลงทุนแบบผสมผสานกันทั้งหลักทรัพย์ที่มีความเสี่ยงต่ำและความเสี่ยงสูง ก็สามารถลดความเสี่ยงลงได้ โดยผู้ลงทุนจะต้องการผลตอบแทน ณ จุด B การกระจายการลงทุนแบบนี้จะสามารถจัดความเสี่ยงส่วนที่เป็นความเสี่ยงเฉพาะตัวของหลักทรัพย์ได้ ความเสี่ยงส่วนที่ยังคงเหลืออยู่ของกลุ่มหลักทรัพย์จะมีเพียงความเสี่ยงที่เป็นระบบที่มีค่าเบต้า (Beta) เป็นตัวชี้วัด ซึ่งการลงทุนในสินทรัพย์ใดๆนั้น ผู้ลงทุนจะพยายามค้นหากลุ่มหลักทรัพย์ที่ให้ผลตอบแทนที่สอดคล้องกับระดับความเสี่ยงที่ผู้ลงทุนสามารถยอมรับได้

การวัดความเสี่ยงของกลุ่มหลักทรัพย์

การวัดความเสี่ยงของกลุ่มหลักทรัพย์ (Risk measure for portfolio) ก่อนอื่นจะต้องทราบแนวคิดพื้นฐานทางสถิติเกี่ยวกับค่าความแปรปรวนร่วมระหว่างอัตราผลตอบแทนของหลักทรัพย์ 2 ตัว ก่อนการคำนวณค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของกลุ่มหลักทรัพย์ ค่าความแปรปรวนร่วมระหว่างอัตราผลตอบแทนของหลักทรัพย์นี้เป็นค่าที่บ่งบอกถึงทิศทางและระดับความผันผวนของอัตราผลตอบแทนของหลักทรัพย์ ว่ามีการเคลื่อนไหวไปในทิศทางเดียวกัน (แสดงค่าบวก) หรือทิศทางตรงกันข้าม (แสดงค่าลบ) และระดับการเคลื่อนไหวไปในทิศทางเดียวกันหรือตรงข้ามนั้นมีมากน้อยเพียงใด

ความแปรปรวนร่วม (Covariance)

การคำนวณค่าความแปรปรวนร่วม เป็นการคำนวณค่าความแปรปรวนร่วมระหว่างอัตราผลตอบแทนของหลักทรัพย์แต่ละคู่ในกลุ่มหลักทรัพย์ หากพิจารณาหลักทรัพย์ A และ B ที่ ประกอบอยู่ในกลุ่มหลักทรัพย์หนึ่ง ความแปรปรวนร่วมระหว่างหลักทรัพย์ A และ B เขียนแทนด้วย สัญลักษณ์ σ_{AB} สามารถคำนวณได้ดังนี้

$$\sigma_{AB} = \sum P_i [R_{Ai} - E(R_A)] [R_{Bi} - E(R_B)]$$

โดยที่ σ_{AB} = ความแปรปรวนร่วมระหว่างอัตราผลตอบแทนของหลักทรัพย์ A และ B

P_i = โอกาสความน่าจะเป็นที่จะเกิดเหตุการณ์ที่ i ในจำนวนเหตุการณ์ m เหตุการณ์

R_{Ai} = อัตราผลตอบแทนที่เป็นไปได้ของหลักทรัพย์ A ตามเหตุการณ์ที่ i

$E(R_A)$ = อัตราผลตอบแทนที่คาดหวังของหลักทรัพย์ A

R_{Bi} = อัตราผลตอบแทนที่เป็นไปได้ของหลักทรัพย์ B ตามเหตุการณ์ที่ i

$E(R_B)$ = อัตราผลตอบแทนที่คาดหวังของหลักทรัพย์ B

ค่าความแปรปรวนร่วมระหว่างอัตราผลตอบแทนของหลักทรัพย์ 2 ชนิด จะเป็นค่าที่บอกถึงทิศทางและระดับความผันผวนของอัตราผลตอบแทนของหลักทรัพย์ 2 ชนิด ว่ามีความผันผวนไปด้วยกัน (เครื่องหมายบวก) หรือสวนทางกัน (เครื่องหมายลบ) และระดับความผันผวนนั้นมีมากน้อยเพียงใด

อัตราผลตอบแทนที่คาดหวังของกลุ่มหลักทรัพย์

อัตราผลตอบแทนที่คาดหวังของกลุ่มหลักทรัพย์ (Expected rate of return of the portfolio) เป็นการคำนวณอัตราผลตอบแทนจากกลุ่มหลักทรัพย์ที่ประกอบด้วยหลักทรัพย์จำนวนมากกว่า 1 หลักทรัพย์ขึ้นไป ที่ประกอบเป็นกลุ่มหลักทรัพย์นั้น โดยอัตราผลตอบแทนที่คาดหวังของกลุ่มหลักทรัพย์สามารถคำนวณได้จากค่าเฉลี่ยถ่วงน้ำหนักของอัตราผลตอบแทนที่คาดหวังของ 8 หลักทรัพย์เดี่ยวแต่ละตัวที่ประกอบเป็นกลุ่ม

หลักทรัพย์ โดยที่อัตราผลตอบแทนของหลักทรัพย์เดี่ยวจะถูกถ่วงน้ำหนักโดยใช้สัดส่วนของเงินลงทุนในหลักทรัพย์นั้นเมื่อเทียบกับเงินลงทุนทั้งหมดของผู้ลงทุน ดังนั้น จากสมการจะเห็นว่า หากกลุ่มหลักทรัพย์ประกอบด้วยหลักทรัพย์ n ตัว เราสามารถคำนวณอัตราผลตอบแทนที่คาดหวังของกลุ่มหลักทรัพย์ได้ ดังสมการ

$$E(R_p) = \sum_{A=1}^n w_A E(R_A)$$

โดยที่ $E(R_p)$ = อัตราผลตอบแทนที่คาดหวังของกลุ่มหลักทรัพย์

w_A = สัดส่วนของเงินลงทุนในหลักทรัพย์ A

$E(R_A)$ = อัตราผลตอบแทนที่คาดหวังของหลักทรัพย์ A

n = จำนวนหลักทรัพย์ในกลุ่มหลักทรัพย์

การคำนวณอัตราผลตอบแทนที่คาดหวังของกลุ่มหลักทรัพย์จะขึ้นอยู่กับสัดส่วนการลงทุนในแต่ละหลักทรัพย์ และอัตราผลตอบแทนของหลักทรัพย์แต่ละตัวที่ประกอบขึ้นเป็นกลุ่มหลักทรัพย์นั้น โดยอัตราผลตอบแทนที่คาดหวังของกลุ่มหลักทรัพย์จะมีค่าสูงขึ้น หากผู้ลงทุนจัดสรรเงินลงทุนในหลักทรัพย์ที่ให้อัตราผลตอบแทนสูงกว่า

การหาสัดส่วนพอร์ตการลงทุนที่เหมาะสมตามทฤษฎี Markowitz

การหาสัดส่วนพอร์ตการลงทุนที่เหมาะสมที่สุดในแต่ละหมวดธุรกิจโดยใช้ Markowitz Mean-Variance Analysis ภายใต้วัตถุประสงค์ที่ทำให้ค่า Sharpe Ratio ของพอร์ตสูงสุดและข้อจำกัดที่ห้ามมีการทำ Short Sales (Beninga, 2008: 336) (Markowitz, 1959) (Parkinson, 2020) สามารถดำเนินการโดยใช้ Solver Function ใน Microsoft Excel ตามขั้นตอนดังต่อไปนี้

1. สร้าง Variance–Covariance Matrix ขนาด 27×27 โดยใช้อัตราผลตอบแทนดัชนีราคาระหว่างหมวดธุรกิจที่เกิดขึ้นจริง โดย Variance–Covariance Matrix ระหว่างหมวดธุรกิจจะประกอบด้วยค่า ความแปรปรวน (Variance) ในแนวทแยง และค่าความแปรปรวนร่วม (Covariance) นอกแนวทแยงดัง Matrix ต่อไปนี้

$$S = \begin{bmatrix} \sigma_{11}^2 & \cdots & \text{cov}_{1j} \\ \vdots & \ddots & \vdots \\ \text{cov}_{ji} & \cdots & \sigma_{jj}^2 \end{bmatrix}$$

2. หาผลตอบแทนคาดหวัง (Expected Return) ของพอร์ตจากสมการดังต่อไปนี้

$$E(r_p) = \sum_{i=1}^n w_i \times E(r_i)$$

โดย $E(r_i)$ คือผลตอบแทนคาดหวัง (Expected Return) จากแต่ละหมวดธุรกิจซึ่งคำนวณจากอัตราผลตอบแทนเฉลี่ยที่เกิดขึ้นจริงในระยะเวลา 5 ปีที่ทำการศึกษา เมื่อนำมาจัดในรูปแบบของ Matrix จะได้ ดังนี้

$$E(r_p) = W^T R = [w_i \cdots w_j] \begin{bmatrix} E(r_1) \\ \vdots \\ E(r_j) \end{bmatrix}$$

โดย W คือเวกเตอร์ของน้ำหนักการลงทุนในแต่ละหมวดธุรกิจ i ถึง j ในพอร์ตและ คือ เวกเตอร์ของผลตอบแทนคาดหวัง (Expected Return)

โดยอัตราผลตอบแทนของหุ้นแต่ละตัวที่สามารถคำนวณได้จากสมการ Geometric mean ดังภาพ

$$GAR = \sqrt[n]{(1 + r_1) * (1 + r_2) * (1 + r_n)} - 1$$

3. คำนวณค่าความเสี่ยง (Standard Deviation) ของพอร์ต โดยเริ่มต้นจากการหา Variance ของพอร์ตที่มีการกระจายลงทุนมากกว่า 2 หมวดธุรกิจจากสมการดังนี้

$$V = \sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \text{Cov}(r_i, r_j)$$

$$\sigma_p^2 = W^T S(W)$$

$$\sigma_p = \sqrt{W^T S(W)} = \left[w_1 \quad \dots \quad w_j \right] \begin{bmatrix} \sigma_{11}^2 & \dots & \text{cov}_{1j} \\ \vdots & \ddots & \vdots \\ \text{cov}_{j1} & \dots & \sigma_{jj}^2 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_j \end{bmatrix} \right]^{\frac{1}{2}}$$

โดย W คือเวกเตอร์ของน้ำหนักการลงทุนในหุ้นแต่ละตัวตั้งแต่ i ถึง j ใน Portfolio และ คือ Variance–Covariance Matrix ของอัตราผลตอบแทนระหว่างหุ้นแต่ละตัว ที่เกิดขึ้นจริง

4. หาค่า Sharpe Ratio ของ Portfolio จากสมการดังต่อไปนี้

$$S_p = \frac{E(r_p) - R_{rf}}{\sigma_p}$$

5. หา Optimal Weighted Portfolio ให้หาค่าที่มี Sharpe Ratio สูงสุด (Maximize) จากนั้นกำหนดให้ Variable Cells เป็นค่าน้ำหนักการลงทุนในแต่ละหุ้น และ ระบุข้อจำกัด (Constraints) โดยผลรวมของน้ำหนักการลงทุนในแต่ละหมวดธุรกิจต้องเท่ากับ 1 และน้ำหนักการลงทุนในแต่ละหมวดธุรกิจจะต้องมีค่ามากกว่าหรือเท่ากับศูนย์

$$\sum_{i=1}^N w_i = 1 \quad \text{โดย } w_i \geq 0$$

ขั้นตอนและวิธีการดำเนินงาน

ข้อมูลที่ใช้ในงานวิจัย

1. ข้อมูลราคาปิดของหุ้น 41 ตัว จาก SET 50

Symbol	Company	Sector
ADVANC	ADVANCED INFO SERVICE PCL	Information & Communication Technology
AOT	AIRPORTS OF THAILAND PCL	Transportation & Logistics
BANPU	BANPU PCL	Energy & Utilities
BBL	BANGKOK BANK PCL	Banking
BDMS	BANGKOK DUSIT MEDICAL SERVICES PCL	Health Care Services
BEM	BANGKOK EXPRESSWAY AND METRO PCL	Transportation & Logistics
BH	BUMRUNGRAD HOSPITAL PCL	Health Care Services
BLA	BANGKOK LIFE ASSURANCE PCL	Finance and Securities
BTS	BTS GROUP HOLDINGS PCL	Transportation & Logistics
CBG	CARABAO GROUP PCL	Food and Beverage
CPALL	CP ALL PCL	Commerce
CPF	CHAROEN POKPHAND FOODS PCL	Food and Beverage
CPN	CENTRAL PATTANA PCL	Property Development
DTAC	TOTAL ACCESS COMMUNICATION PCL	Information & Communication Technology
EA	ENERGY ABSOLUTE PCL	Energy & Utilities
EGCO	ELECTRICITY GENERATING PCL	Energy & Utilities
GLOBAL	SIAM GLOBAL HOUSE PCL	Commerce
GPSC	GLOBAL POWER SYNERGY PCL	Energy & Utilities
HMPRO	HOME PRODUCT CENTER PCL	Commerce
INTUCH	INTOUCH HOLDINGS PCL	Information & Communication Technology
IRPC	IRPC PCL	Energy & Utilities
IVL	INDORAMA VENTURES PCL	Petrochemicals & Chemicals
JMART	JAY MART PCL	Information & Communication Technology
JMT	JMT NETWORK SERVICES PCL	Finance and Securities
KBANK	KASIKORNBANK PCL	Banking
KCE	KCE ELECTRONICS PCL	Electronic Components
KTB	KRUNG THAI BANK PCL	Banking

KTC	KRUNGTHAI CARD PCL	Finance and Securities
LH	LAND AND HOUSES PCL	Property Development
MINT	MINOR INTERNATIONAL PCL	Food and Beverage
MTC	MUANGTHAI CAPITAL PCL	Finance and Securities
PTT	PTT PCL	Energy & Utilities
PTTEP	PTT EXPLORATION AND PRODUCTION PCL	Energy & Utilities
PTTGC	PTT GLOBAL CHEMICAL PCL	Petrochemicals & Chemicals
SAWAD	SRISAWAD CORPORATION PCL	Finance and Securities
SCC	THE SIAM CEMENT PCL	Construction Materials
TISCO	TIPCO ASPHALT PCL	Construction Materials
TOP	THAI OIL PCL	Energy & Utilities
TRUE	TRUE CORPORATION PCL	Information & Communication Technology
TTB	TMBTHANACHART BANK PCL	Banking
TU	THAI UNION GROUP PCL	Food and Beverage

2.ข้อมูล financial ratio 5 ตัว ได้แก่

- อัตราส่วนหนี้สินต่อทุน (Debt to Equity Ratio)
- อัตราส่วนหมุนเวียนสินทรัพย์ (Assets Turnover Ratio)
- อัตราส่วนผลตอบแทนต่อสินทรัพย์รวม (Return on Assets)
- ราคาหุ้นต่อมูลค่าหุ้นทางบัญชี (Price to Book Value)
- อัตราส่วนเงินปันผลตอบแทน (Dividend Yield)

3.ข้อมูลราคาปิดของ SET index

ขั้นตอนการวิจัย

ในการวิจัยครั้งนี้ได้แบ่งขั้นตอนการวิจัยเป็น 3 ส่วน คือ

1) การจัดห้เป็นกลุ่ม 7 กลุ่มด้วยวิธี k-means clustering

ขั้นตอนที่ 1 : ติดตั้ง library package

```
##### Library packages #####
library(tidyverse) # data manipulation
library(tidyquant) # for finance data
library(factoextra) # clustering visualization for k-means
library(dendextend) # for comparing two dendrograms for k-means
```

ขั้นตอนที่ 2 : สร้าง return_data ที่มีข้อมูลอัตราผลตอบแทนของหุ้น(stock_ret) 41 ตัว และ อัตราผลตอบแทนของตลาด(set_index_return)

ดึงข้อมูลราคาซื้อขายหุ้น 41 ตัว จาก yahoo finance และเลือกเฉพาะคอลัมน์ adjusted ซึ่งเป็นราคาปิดของหุ้นในแต่ละวันที่คิดผลของการจ่ายเงินปันผลแล้ว มาหาอัตราผลตอบแทนรายวัน(stock_ret)

```

# Set date
date_from_kmeans = "2021-01-01"
date_to_kmeans = "2021-12-31"

# Create name of stock in SET50 (exclUde AWC.BK, BRGRIM.BK, CRC.BK, GULF.BK,
# ,OSP.BK, SCGP.BK, SCB.BK, TIDLOR.BK and OR.BK)
# summary 41 symbols
symbols_set50 <- c("ADVANC.BK", "AOT.BK", "BANPU.BK", "BBL.BK", "BDMS.BK",
"BEM.BK", "BH.BK", "BLA.BK", "BTS.BK", "CBG.BK",
"CPALL.BK", "CPF.BK", "CPN.BK", "DTAC.BK", "EA.BK",
"EGCO.BK", "GLOBAL.BK", "GPSC.BK", "HMPRO.BK", "INTUCH.BK",
"IRPC.BK", "IVL.BK", "JMART.BK", "JMT.BK", "KBANK.BK",
"KCE.BK", "KTB.BK", "KTC.BK", "LH.BK", "MINT.BK",
"MTB.BK", "PTT.BK", "PTTEP.BK", "PTTGC.BK", "SAWAD.BK",
"SCC.BK", "TISCO.BK", "TOP.BK", "TRUE.BK", "TTB.BK",
"TTU.BK")

```

	symbol	date	stock_ret
1	ADVANC.BK	2021-01-04	0.0000000000
2	ADVANC.BK	2021-01-05	0.0142044093
3	ADVANC.BK	2021-01-06	0.0056023511
4	ADVANC.BK	2021-01-07	0.0055710495
5	ADVANC.BK	2021-01-08	0.0221605601
6	ADVANC.BK	2021-01-11	-0.0135501825
7	ADVANC.BK	2021-01-12	-0.0082416094
8	ADVANC.BK	2021-01-13	0.0055400952
9	ADVANC.BK	2021-01-14	-0.0055095717
10	ADVANC.BK	2021-01-15	-0.0027700924

ดึงข้อมูลราคาซื้อขายหุ้นของ SET จาก yahoo finance และเลือกเฉพาะคอลัมน์ adjusted ซึ่งเป็นราคาปิดของหุ้นในแต่ละวันที่คิดผลของการจ่ายเงินปันผลแล้ว มาหาอัตราผลตอบแทนรายวัน (set_index_return)

	date	set_ret
1	2021-01-04	0.000000e+00
2	2021-01-05	2.616060e-02
3	2021-01-06	-9.484644e-03
4	2021-01-07	1.435313e-02
5	2021-01-08	1.496909e-02
6	2021-01-11	3.257465e-05
7	2021-01-12	2.186793e-03
8	2021-01-13	4.844682e-03
9	2021-01-14	-7.322436e-03
10	2021-01-15	-1.097018e-02

ใช้ fuction inner_join รวมตาราง stocks_return กับ ตาราง set_index_return

	symbol	date	stock_ret	set_ret
1	ADVANC.BK	2021-01-04	0.0000000000	0.000000e+00
2	ADVANC.BK	2021-01-05	0.0142044093	2.616060e-02
3	ADVANC.BK	2021-01-06	0.0056023511	-9.484644e-03
4	ADVANC.BK	2021-01-07	0.0055710495	1.435313e-02
5	ADVANC.BK	2021-01-08	0.0221605601	1.496909e-02
6	ADVANC.BK	2021-01-11	-0.0135501825	3.257465e-05
7	ADVANC.BK	2021-01-12	-0.0082416094	2.186793e-03
8	ADVANC.BK	2021-01-13	0.0055400952	4.844682e-03
9	ADVANC.BK	2021-01-14	-0.0055095717	-7.322436e-03
10	ADVANC.BK	2021-01-15	-0.0027700924	-1.097018e-02

ขั้นตอนที่ 3 : หาค่า beta จากสมการ capm : $\text{return} = \text{risk free} + \text{beta} \times (\text{market return} - \text{risk free})$

นำตาราง return_data ที่มีอัตราผลตอบแทนรายวันของหุ้นของหุ้น(return) 41 ตัว กับอัตราผลตอบแทนของหุ้นใน SET (market return)

	symbol	Beta
1	ADVANC.BK	0.5921
2	AOT.BK	1.3492
3	BANPU.BK	0.7524
4	BBL.BK	1.2572
5	BDMS.BK	0.6008
6	BEM.BK	0.9207
7	BH.BK	0.7972
8	BLA.BK	0.6026
9	BTS.BK	1.0127
10	CBG.BK	0.6249

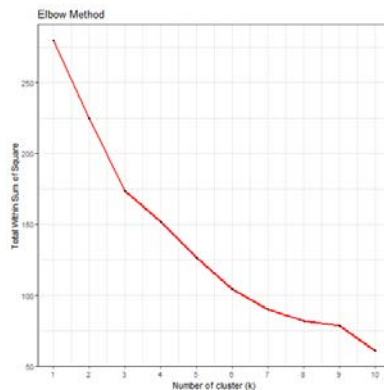
ขั้นตอนที่ 4 : สร้างตาราง dataset ที่มีข้อมูลอัตราผลตอบแทนเฉลี่ยของหุ้น 41 ตัว ค่า Beta และ financial ratio 5 ตัว และ คำนวณค่าเฉลี่ยอัตราผลตอบแทนของหุ้น 41 ตัว แล้ว inner_join กับค่า Beta และ financial ratio 5 ตัว

	symbol	mean_return	Beta	debt_to_equity	asset_turnover	roe	pbv	dvd_yield
1	ADVANC.BK	1.349350e-03	0.5921	3.36	0.52	34.24	9.15	3.01
2	AOT.BK	7.306476e-05	1.3492	0.73	0.04	-12.82	7.76	0.00
3	BANPU.BK	3.877768e-04	0.7524	3.31	0.42	13.91	1.04	2.12
4	BBL.BK	2.649335e-04	1.2572	7.79	0.04	5.63	0.48	2.06
5	BDMS.BK	6.430962e-04	0.6008	0.49	0.57	9.24	4.53	2.39
6	BEM.BK	1.884942e-04	0.9207	2.05	0.10	2.67	3.41	1.18
7	BH.BK	8.842257e-04	0.7972	0.19	0.56	6.79	6.75	2.27
8	BLA.BK	2.499276e-03	0.6026	6.24	0.14	6.80	1.34	0.62
9	BTS.BK	2.364125e-04	1.0127	2.26	0.22	7.95	2.06	3.31
10	CBG.BK	3.989419e-04	0.6249	0.89	0.97	28.57	12.44	2.01

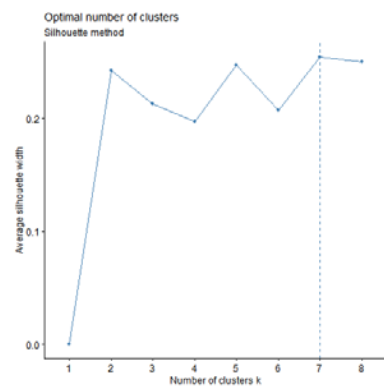
ขั้นตอนที่ 5 : ทำให้ข้อมูลในตาราง dataset โดยใช้วิธี z-score ทำให้ข้อมูลมีค่าเฉลี่ยเท่ากับ 1 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0

	mean_return	Beta	debt_to_equity	asset_turnover	roe	pbv	dvd_yield
ADVANC.BK	0.36630462	-1.16525513	0.23572668	0.21790727	2.15496497	1.3195809	0.49217449
AOT.BK	-0.82169771	0.98620637	-0.83723271	-1.09755745	-2.48904194	0.9541635	-1.44713758
BANPU.BK	-0.52875494	-0.70972841	0.21532821	-0.05614788	0.14874609	-0.8124592	-0.08124336
BBL.BK	-0.64310097	0.72476870	2.04303090	-1.09755745	-0.66834650	-0.9596777	-0.11990075
BDMS.BK	-0.29109656	-1.14053222	-0.93514535	0.35493484	-0.31210203	0.1050279	0.09271486
BEM.BK	-0.71425279	-0.23146798	-0.29871317	-0.93312436	-0.96044723	-0.1894092	-0.68687570
BH.BK	-0.06664646	-0.58241964	-1.05753616	0.32752932	-0.55387459	0.6886443	0.01540010
BLA.BK	1.43668727	-1.13541714	1.41067841	-0.82350230	-0.55288777	-0.7335921	-1.04767795
BTS.BK	-0.66964911	0.02996968	-0.21303960	-0.60425818	-0.43940268	-0.5443111	0.68546141
CBG.BK	-0.51836214	-1.07204692	-0.77195761	1.45115544	1.59543417	2.1844899	-0.15211523
CPALL.BK	-0.69264030	-0.25278084	1.36988147	1.01266720	0.05203706	0.3652892	-0.46137430

ขั้นตอนที่ 6 : หาค่า K หรือจำนวนกลุ่มที่เหมาะสม โดยใช้วิธี Elbow และ Silhouette Method
วิธี Elbow Method เมื่อพิจารณาจากกราฟจะพบว่า k=3 ดีที่สุด แต่ไม่ชัดเจนมาก



วิธี Silhouette Method ให้ค่า k=7 ซึ่งวิธีนี้จะเป็นวิธีที่เลือกค่า k มาให้โดยที่เราต้องมองหาค่า k จากกราฟเอง ดังนั้น เราจึงเลือกค่า K=7 หรือเลือกจำนวนกลุ่มที่จะแบ่งเท่ากับ 7 กลุ่ม



ขั้นตอนที่ 7 : นำตาราง dataset จากขั้นตอนที่ 5 มาเข้า k-means cluster model เพื่อแบ่งกลุ่มหุ้นออกมาเป็น 7 กลุ่ม และ set seed เพื่อให้กลุ่มที่สุ่มออกมาได้ค่าที่สุ่มชุดเดิมทุกครั้ง และไม่ให้เลขของกลุ่มเปลี่ยนไปเปลี่ยนมาและใช้ฟังก์ชัน kmeans ของ library stat เพื่อจัดกลุ่มข้อมูลทั้ง 7

```
K-means clustering with 7 clusters of sizes 7, 3, 6, 14, 3, 2, 6

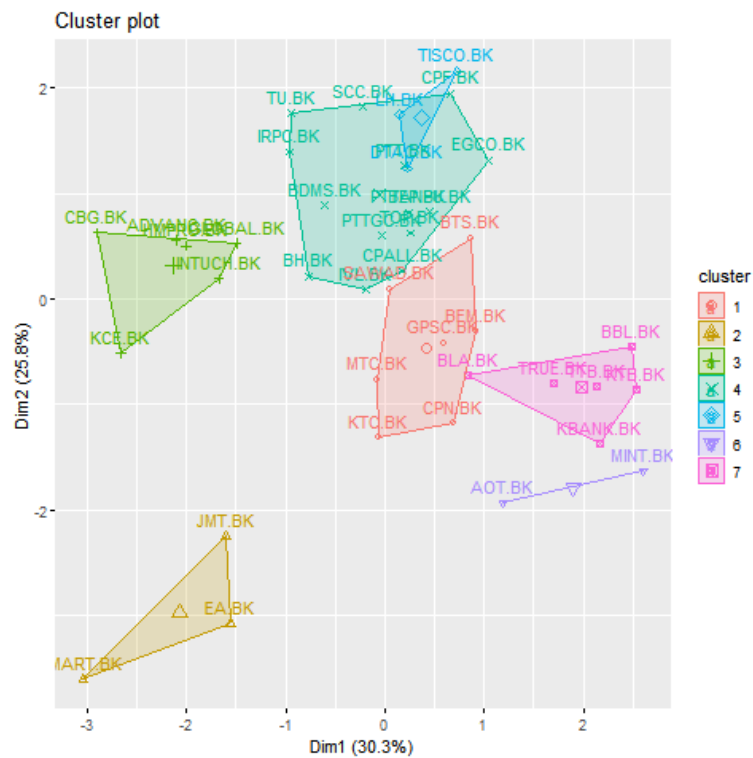
Cluster means:
  mean return      Beta debt to equity asset turnover      roe      pbv      dvd_yield
1 -0.5600045    0.7216428    -0.2952163    -0.6708144    0.09024136    -0.1019043    -0.3012223
2  2.3587212    1.1218508    -0.7148419    -0.5951230    0.58656824    2.2291813    -1.0090206
3  0.4078445    -0.9627830    -0.6169293    0.7842879    1.27603123    1.1272329    -0.1574843
4 -0.4987578    -0.3654954    -0.3765188    0.7660176    -0.06377411    -0.5411189    0.1994829
5  0.1253687    -0.9499480    0.7076112    -0.5220416    0.28262559    -0.2481214    2.5861161
6 -0.6625458    1.0069509    -0.1151270    -0.8646106    -2.81913551    0.2995667    -1.4471376
7  0.3880659    0.5520872    1.8818830    -0.9422595    -0.72739164    -0.8361193    -0.2627183

Clustering vector:
ADVANC.BK  AOT.BK  BANPU.BK  BBL.BK  BDM5.BK  BEM.BK  BH.BK  BLA.BK  BTS.BK  CBG.BK  CPALL.BK
3          6          4          7          4          1          4          7          1          3          4
CPF.BK    CPN.BK  DTAC.BK  EA.BK  EGCO.BK  GLOBAL.BK  GPSC.BK  HMPRO.BK  INTUCH.BK  IRPC.BK  IVL.BK
4          1          5          2          4          3          1          3          3          4          4
JMAT.BK   JMT.BK  KBANK.BK  KCE.BK  KTB.BK  KTC.BK  LH.BK  MINT.BK  MITC.BK  PTT.BK  PTTEP.BK
2          2          7          3          7          1          5          6          1          4          4
PTTGC.BK  SAMAD.BK  SCC.BK  TISCO.BK  TOP.BK  TRUE.BK  TTB.BK  TU.BK
4          1          4          5          4          7          7          4

Within cluster sum of squares by cluster:
[1] 12.062580  4.780205 18.876349 36.469371  5.032415  2.277839 10.392982
(between_SS / total_SS = 67.9 %)

Available components:
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"    "size"
[8] "iter"         "ifault"
```

ใช้ฟังก์ชัน fviz_cluster() จาก library factoextra โดยได้เ้ามาจากการใช้ ggplot2 ในการสร้างกราฟ



ขั้นตอนที่ 8 : เก็บเลขที่จัดกลุ่มของหุ้น (km.cluster) และที่มีข้อมูล ratio ต่างๆของแต่ละหุ้นไว้ใน dataframe ที่เป็น cluster

	symbol	mean_return	Beta	debt_to_equity	asset_turnover	roe	pbv	dvd_yield	km.cluster
1	ADVANC.BK	1.349350e-03	0.5921	3.36	0.52	34.24	9.15	3.01	3
2	AOT.BK	7.306476e-05	1.3492	0.73	0.04	-12.82	7.76	0.00	6
3	BANPU.BK	3.877768e-04	0.7524	3.31	0.42	13.91	1.04	2.12	4
4	BBL.BK	2.649335e-04	1.2572	7.79	0.04	5.63	0.48	2.06	7
5	BDMS.BK	6.430962e-04	0.6008	0.49	0.57	9.24	4.53	2.39	4
6	BEM.BK	1.884942e-04	0.9207	2.05	0.10	2.67	3.41	1.18	1
7	BH.BK	8.842257e-04	0.7972	0.19	0.56	6.79	6.75	2.27	4
8	BLA.BK	2.499276e-03	0.6026	6.24	0.14	6.80	1.34	0.62	7
9	BTS.BK	2.364125e-04	1.0127	2.26	0.22	7.95	2.06	3.31	1
10	CBG.BK	3.989419e-04	0.6249	0.89	0.97	28.57	12.44	2.01	3
11	CPALL.BK	2.117127e-04	0.9132	6.14	0.81	12.93	5.52	1.53	4

2) การหาหน้าหนักที่เหมาะสมของ portfolio ตามทฤษฎีพอร์ตการลงทุนสมัยใหม่ (Modern Portfolio Theory) ของ Markowitz

ขั้นตอนที่ 1 : ติดตั้ง library packages

```
library(tidyverse) # data manipulation
library(tidyquant) # for finance data
library(plotly) # To create interactive charts
library(timetk) # To manipulate the data series
```

ขั้นตอนที่ 2 : ตั้งวันที่ให้ใช้ข้อมูล 5 ปี สำหรับ Optimization ตั้งแต่ปี 2016 ถึง 2021 และตั้งตัวแปรที่เก็บชื่อหุ้นทั้ง 41 ตัว

```
# Set date
date_from = "2016-01-01"
date_to = "2021-12-31"

# Create name of stock in SET50 (exclude AWC.BK, BRGRIM.BK, CRC.BK, GULF.BK,
# ,OSP.BK, SCGP.BK, SCB.BK, TIDLOR.BK and OR.BK)
# summary 41 symbols
symbols_set50 <- c("ADVANC.BK", "AOT.BK", "BANPU.BK", "BBL.BK", "BDMS.BK",
  "BEM.BK", "BH.BK", "BLA.BK", "BTS.BK", "CBG.BK",
  "CPALL.BK", "CPF.BK", "CPN.BK", "DTAC.BK", "EA.BK",
  "EGCO.BK", "GLOBAL.BK", "GPSC.BK", "HMPRO.BK", "INTUCH.BK",
  "IRPC.BK", "IVL.BK", "JMART.BK", "JMT.BK", "KBANK.BK",
  "KCE.BK", "KTB.BK", "KTC.BK", "LH.BK", "MINT.BK",
  "MTC.BK", "PTT.BK", "PTTEP.BK", "PTTGC.BK", "SAWAD.BK",
  "SCC.BK", "TISCO.BK", "TOP.BK", "TRUE.BK", "TTB.BK",
  "TU.BK")
```

ขั้นตอนที่ 3 : เลือกหุ้นที่ดีที่สุดจากทั้ง 7 กลุ่ม หรือ 7 cluster ที่ได้ k-means clustering และนำแต่ละกลุ่มเหล่านั้นมาหาค่า Sharpe Ratio ที่มากที่สุดของหุ้นแต่ละตัวในหุ้นแต่ละกลุ่ม ดังนั้นจะได้หุ้น 7 ตัวที่มีค่า Sharpe Ratio มากที่สุดของแต่ละกลุ่มมาจัด portfolio

	symbol	sharpe_ratio
1	GPSC.BK	0.91114445
2	JMT.BK	1.35010755
3	GLOBAL.BK	0.62946099
4	PTTEP.BK	0.52478995
5	TISCO.BK	0.93799036
6	AOT.BK	0.41185046
7	KTB.BK	0.08869614

ขั้นตอนที่ 4 : สร้างฟังก์ชันสำหรับการหา optimal portfolio ด้วยการสุ่มน้ำหนัก 100,000 รอบ โดยภายในฟังก์ชันจะมีขั้นตอนวิธีการทำดังต่อไปนี้

1. ดึงราคาจาก yahoo finance และนำมาคำนวณอัตราผลตอบแทน (Return) ของหุ้นแต่ละตัวโดยกำหนดให้ใช้ “daily” เพื่อให้คำนวณอัตราผลตอบแทนรายวัน โดยที่ใช้ข้อมูลราคาที่ adjusted แล้ว
2. นำข้อมูลอัตราผลตอบแทนรายวันของหุ้นแต่ละตัวมาคำนวณ Geometric Return ของหุ้นแต่ละตัว เนื่องจากเป็นข้อมูลอัตราผลตอบแทนรายวันดังนั้นจะให้ Annualized Return ด้วย 252 วัน เพื่อให้ให้อัตราผลตอบแทนอยู่ในหน่วยรายปี

	symbol	AnnualizedReturn
1	GPSC.BK	0.33299233
2	JMT.BK	0.57832803
3	GLOBAL.BK	0.22122867
4	PTTEP.BK	0.19331310
5	TISCO.BK	0.23720778
6	AOT.BK	0.11750486
7	KTB.BK	0.02216795

3. คำนวณ Covariance Matrix ระหว่างหุ้นแต่ละตัวใน Portfolio

	AOT.BK	GLOBAL.BK	GPSC.BK	JMT.BK	KTB.BK	PTTEP.BK	TISCO.BK
AOT.BK	0.08140157	0.02869625	0.03789574	0.02714438	0.03424002	0.04398976	0.02660555
GLOBAL.BK	0.02869625	0.12352222	0.03435445	0.04435113	0.02518193	0.03115597	0.02172752
GPSC.BK	0.03789574	0.03435445	0.13356546	0.04035892	0.02735997	0.04093419	0.02213127
JMT.BK	0.02714438	0.04435113	0.04035892	0.18348966	0.02431935	0.02756682	0.01880395
KTB.BK	0.03424002	0.02518193	0.02735997	0.02431935	0.06246569	0.04471312	0.02706909
PTTEP.BK	0.04398976	0.03115597	0.04093419	0.02756682	0.04471312	0.13569120	0.02884749
TISCO.BK	0.02660555	0.02172752	0.02213127	0.01880395	0.02706909	0.02884749	0.06395301

4. กำหนดตัวแปรต่างๆและตัวแปรสำหรับเก็บค่าระหว่างการ Loop

- กำหนดจำนวนครั้งในการ Loop ในการวิจัยครั้งนี้จะใช้ทั้งหมด 100,000 รอบ
- สร้าง Matrix สำหรับเก็บน้ำหนักจากการลงทุนที่ได้จากการสุ่มแต่ละครั้ง
- สร้าง Vector ที่มีข้อมูลประเภท numeric สำหรับบันทึกอัตราผลตอบแทนของ portfolio
- สร้าง Vector ที่มีข้อมูลประเภท numeric สำหรับบันทึกความเสี่ยงของ portfolio
- สร้าง Vector ที่มีข้อมูลประเภท numeric สำหรับบันทึก Sharpe ratio ของ portfolio

5. ใช้ for loop สำหรับการคำนวณอัตราผลตอบแทนของ portfolio, ความเสี่ยงที่วัดด้วย Standard deviation ของ portfolio และ Sharpe Ratio (กำหนดให้ Risk free rate = 0%) ภายใน Loop มีสิ่งที่เกิดขึ้นดังต่อไปนี้

- สุ่มน้ำหนักในการลงทุนขึ้นมาตามจำนวนหุ้น และปรับค่าให้ผลรวมของน้ำหนักในการลงทุนเท่ากับ 1 และบันทึกค่าในกัน Loop แต่ละครั้งไว้ใน Matrix ที่เก็บน้ำหนักในการลงทุน
- คำนวณอัตราผลตอบแทนของ portfolio ด้วยผลรวมของน้ำหนักแต่ละหุ้นและ Geometric mean return ที่คำนวณมาได้มาคูณกันและหาผลรวมของค่าดังกล่าว และบันทึกค่าในกัน Loop แต่ละครั้งไว้ใน Vector ที่เก็บอัตราผลตอบแทนของ portfolio
- คำนวณความเสี่ยงของ portfolio ด้วยผลรวมของน้ำหนักแต่ละหุ้นและ Covariance Matrix ที่คำนวณมา และบันทึกค่าในกัน Loop แต่ละครั้งไว้ใน Vector ที่เก็บความเสี่ยงของ portfolio
- คำนวณ Sharpe Ratio ของ portfolio ด้วยอัตราผลตอบแทนของ portfolio หาดด้วยความเสี่ยงของ portfolio และบันทึกค่าในกัน Loop แต่ละครั้งไว้ใน Vector ที่เก็บ Sharpe Ratio ของ portfolio

6. นำน้ำหนักการลงทุนในแต่ละหุ้น, อัตราผลตอบแทนของ portfolio, ความเสี่ยงที่วัดด้วย Standard deviation ของ portfolio และ Sharpe Ratio มาสร้างตารางเพื่อนบันทึก portfolio ทั้ง 100,000 portfolio

7. แถวที่ทำให้ Sharpe Ratio มากที่สุด และหาแถวที่ทำให้ความเสี่ยงน้อยที่สุดจากข้อมูลในข้อที่ 6

8. สร้างกราฟ bar plot เพื่อแสดงน้ำหนักของการลงทุนที่ทำให้ Sharpe Ratio มากที่สุด และน้ำหนักทำให้ความเสี่ยงน้อยที่สุด

9. สร้างกราฟ Efficient frontier ที่เป็นกราฟจุดที่มีแกน x คือ ความเสี่ยงของ portfolio และแกน y คือ อัตราผลตอบแทนของ portfolio และบนกราฟจะแสดงถึงว่าจุดไหนที่ให้ค่าความเสี่ยงน้อยที่สุด และจุดไหนจะให้ค่า Sharpe Ratio มากที่สุด

10. ส่งค่ากลับไปยังฟังก์ชัน โดยมีค่าดังต่อไปนี้

- ตารางที่เก็บน้ำหนักที่ไว้สร้าง portfolio 10000 แถว
- ตารางที่เก็บ portfolio ที่มีชื่อหุ้น, น้ำหนักในการลงทุนของหุ้นแต่ละตัว, อัตราผลตอบแทนของ portfolio, ความเสี่ยงของ portfolio และ Sharpe Ratio ของ portfolio ที่ทำให้ความเสี่ยงของ portfolio น้อยที่สุด
- ตารางที่เก็บ portfolio ที่มีชื่อหุ้น, น้ำหนักในการลงทุนของหุ้นแต่ละตัว, อัตราผลตอบแทนของ portfolio, ความเสี่ยงของ portfolio และ Sharpe Ratio ของ portfolio ที่ทำให้ Sharpe Ratio portfolio มากที่สุด
- bar plot ที่แสดงถึงน้ำหนักที่ทำให้ความเสี่ยงของ portfolio น้อยที่สุด
- bar plot ที่แสดงถึงน้ำหนักที่ทำให้ Sharpe Ratio ของ portfolio มากที่สุด
- Efficient frontier

ขั้นตอนที่ 5 : นำผลที่ได้จากฟังก์ชันมาแสดงและบันทึกค่าเหล่านั้นไว้ถ้าเป็นตารางเก็บเป็นไฟล์นามสกุล csv และถ้าเป็นรูปภาพให้เก็บเป็นไฟล์นามสกุล png สามารถแสดงผลได้ดังต่อไปนี้

เรียกใช้ฟังก์ชันเพื่อรับค่าต่างๆ ดังต่อไปนี้

- ตารางที่ชื่อ portfolio_values ที่เก็บ portfolio ไว้ทั้งหมด 100,000 portfolio

	GPSC.BK	JMT.BK	GLOBAL.BK	PTTEP.BK	TISCO.BK	AOT.BK	KTB.BK	Return	Risk	SharpeRatio
1	0.149969032	2.258023e-01	0.111152162	0.1654660255	0.025067202	0.261514376	0.0610288880	0.2751313	0.2271238	1.2113716
2	0.148470806	1.191120e-01	0.043650852	0.1650202551	0.122784249	0.202711358	0.1982504458	0.2172226	0.2083281	1.0426945
3	0.160322093	1.422220e-01	0.165624378	0.1405780763	0.145281682	0.146548703	0.0994230523	0.2533396	0.2103325	1.2044716
4	0.284197222	2.513101e-01	0.253133311	0.0320740933	0.051446182	0.003123071	0.1247160586	0.3175109	0.2188482	1.4508270
5	0.217041433	4.462902e-02	0.108923960	0.1255190159	0.103692612	0.232761658	0.1674323002	0.2021039	0.2124611	0.9512512
6	0.207728846	1.264470e-01	0.206566786	0.1782958898	0.153238314	0.091379266	0.0363439206	0.2703579	0.2172901	1.2442255
7	0.251439604	9.002064e-02	0.140226854	0.3520728300	0.092373792	0.062801414	0.0110648714	0.2644079	0.2409097	1.0975397
8	0.187902396	5.821486e-02	0.057927362	0.0971127509	0.212885309	0.226254058	0.1597032675	0.2084500	0.2088678	0.9979998
9	0.178077909	2.748526e-01	0.267595305	0.1023386655	0.054635618	0.106786838	0.0157130819	0.3230930	0.2291377	1.4100383
10	0.198468395	1.998171e-01	0.162326440	0.0487621077	0.166479176	0.120427670	0.1037190987	0.2829261	0.2077831	1.3616415

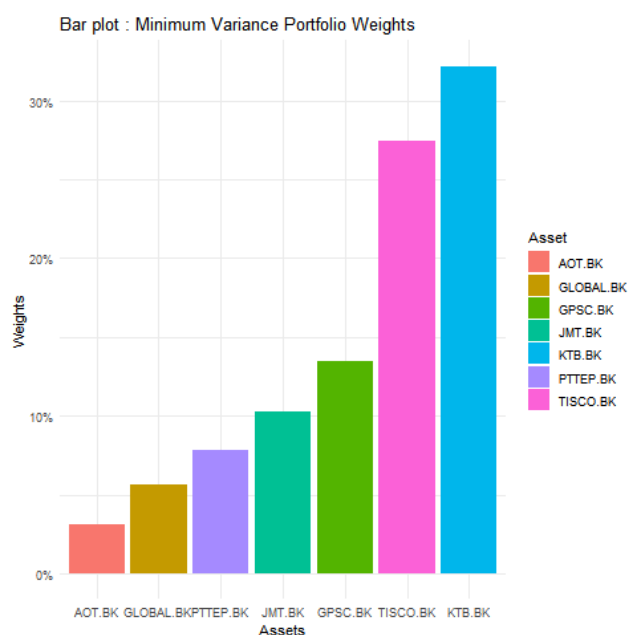
- ตารางที่เก็บ portfolio ที่มีชื่อหุ้น, น้ำหนักในการลงทุนของหุ้นแต่ละตัว, อัตราผลตอบแทนของ portfolio, ความเสี่ยงของ portfolio และ Sharpe Ratio ของ portfolio ที่ทำให้ความเสี่ยงของ portfolio น้อยที่สุด

	GPSC.BK	JMT.BK	GLOBAL.BK	PTTEP.BK	TISCO.BK	AOT.BK	KTB.BK	Return	Risk	SharpeRatio
1	0.1346931	0.1030152	0.0565348	0.07815277	0.2744426	0.03130048	0.321861	0.2079563	0.1937624	1.073254

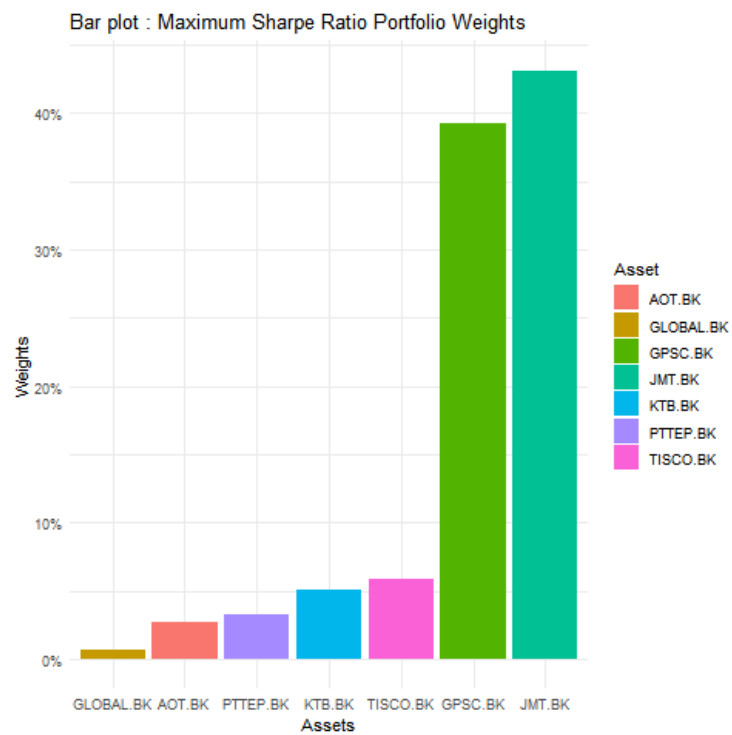
- ตารางที่เก็บ portfolio ที่มีชื่อหุ้น, น้ำหนักในการลงทุนของหุ้นแต่ละตัว, อัตราผลตอบแทนของ portfolio, ความเสี่ยงของ portfolio และ Sharpe Ratio ของ portfolio ที่ทำให้ Sharpe Ratio portfolio มากที่สุด

	GPSC.BK	JMT.BK	GLOBAL.BK	PTTEP.BK	TISCO.BK	AOT.BK	KTB.BK	Return	Risk	SharpeRatio
1	0.3925838	0.4311152	0.007421369	0.03237471	0.05868176	0.02743811	0.05038511	0.4062144	0.2355021	1.724887

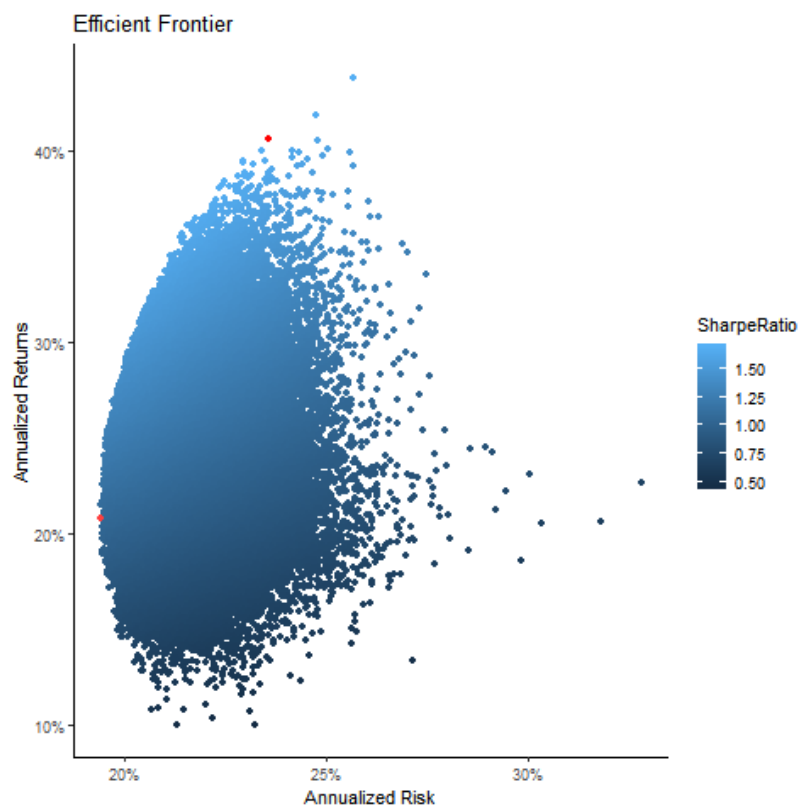
- bar plot ที่แสดงถึงน้ำหนักที่ทำให้ความเสี่ยงของ portfolio น้อยที่สุด



- bar plot ที่แสดงถึงน้ำหนักที่ทำให้ Sharpe Ratio ของ portfolio มากที่สุด



- กราฟ Efficient frontier



3) การเปรียบเทียบอัตราผลตอบแทนสะสม (cumulative return) ของ portfolio ที่มีน้ำหนักที่เหมาะสมตามทฤษฎีพอร์ตการลงทุนสมัยใหม่ กับที่มีน้ำหนักการลงทุนเท่ากัน (Equally weight portfolio) เทียบกับอัตราผลตอบแทนสะสม (cumulative return) ของ SET index

ขั้นตอนที่ 1 : ติดตั้ง library packages และตั้งค่าวันที่จะใช้ข้อมูล 5 ปี ตั้งแต่ปี 2016 ถึง 2021

```
library(tidyverse) # data manipulation
library(tidyquant) # for finance data

# Set date
date_from = "2016-01-01"
date_to = "2021-12-31"
```

ขั้นตอนที่ 2 : ให้นำเข้าข้อมูลที่มี

- ตารางที่เก็บ portfolio ที่มีชื่อหุ้น, น้ำหนักในการลงทุนของหุ้นแต่ละตัว, อัตราผลตอบแทนของ portfolio, ความเสี่ยงของ portfolio และ Sharpe Ratio ของ portfolio ที่ทำให้ความเสี่ยงของ portfolio น้อยที่สุด

	GPSC.BK	JMT.BK	GLOBAL.BK	PTTEP.BK	TISCO.BK	AOT.BK	KTB.BK	Return	Risk	SharpeRatio
1	0.1346931	0.1030152	0.0565348	0.07815277	0.2744426	0.03130048	0.321861	0.2079563	0.1937624	1.073254

- ตารางที่เก็บ portfolio ที่มีชื่อหุ้น, น้ำหนักในการลงทุนของหุ้นแต่ละตัว, อัตราผลตอบแทนของ portfolio, ความเสี่ยงของ portfolio และ Sharpe Ratio ของ portfolio ที่ทำให้ Sharpe Ratio portfolio มากที่สุด

	GPSC.BK	JMT.BK	GLOBAL.BK	PTTEP.BK	TISCO.BK	AOT.BK	KTB.BK	Return	Risk	SharpeRatio
1	0.3925838	0.4311152	0.007421369	0.03237471	0.05868176	0.02743811	0.05038511	0.4062144	0.2355021	1.724887

- vector ของชื่อหุ้นทั้ง 7 ตัวที่ใช้ในการ optimization

```
> portfolio_symbol
[1] "GPSC.BK" "JMT.BK" "GLOBAL.BK" "PTTEP.BK" "TISCO.BK" "AOT.BK" "KTB.BK"
```

ขั้นตอนที่ 3 : สร้างฟังก์ชันไว้สำหรับคำนวณ Optimal weight Portfolio cumulative growth เทียบกับ SET Index Portfolio cumulative growth ขั้นตอนที่เกิดขึ้นภายในฟังก์ชันมีดังนี้

1. ฟังก์ชันนี้จะรับชื่อหุ้นแต่ละตัวที่ต้องการ และ น้ำหนักในการลงทุนของหุ้นแต่ละตัว
2. คำนวณ Portfolio Cumulative growth จากน้ำหนักการลงทุนและหุ้นที่รับเข้ามา โดยใช้ข้อมูลเป็นรายวัน และข้อมูลราคาที่ adjusted
3. คำนวณ SET Index Portfolio cumulative growth โดยใช้ข้อมูลเป็นรายวัน และข้อมูลราคาที่ adjusted
4. สร้าง line graph ที่แสดงถึงความสัมพันธ์ระหว่าง Cumulative return และวันเวลา โดยจะเป็นกราฟลักษณะของ Portfolio ที่ฟังก์ชันรับเข้ามาจากฟังก์ชัน เทียบกับ SET Index ที่ให้เป็นตัวแทนของตลาด
5. ส่ง line graph กลับไปยังจุดที่เรียกใช้ฟังก์ชัน

ขั้นตอนที่ 4 : สร้างฟังก์ชันสำหรับนำเข้าค่าน้ำหนักในการลงทุน และเรียกใช้ฟังก์ชัน ได้ผลดังนี้

น้ำหนักการลงทุนที่ทำให้ความเสี่ยงของ portfolio น้อยที่สุด

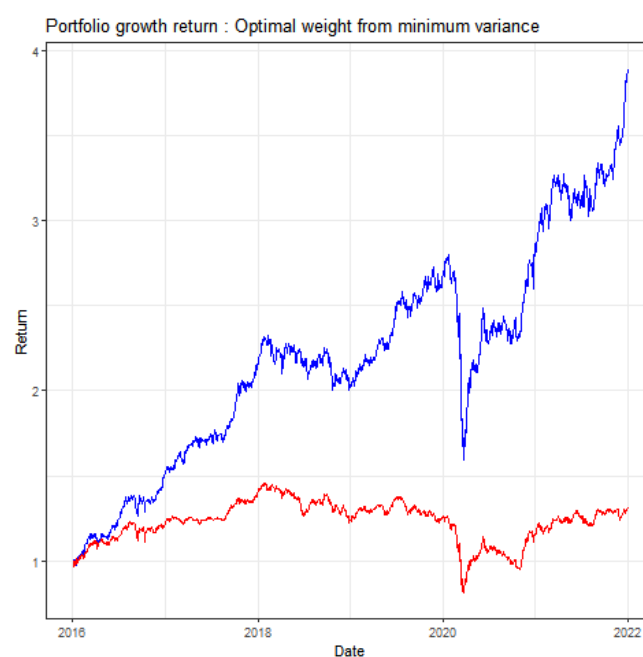
	symbol	weights
1	GPSC.BK	0.11225883
2	JMT.BK	0.10214355
3	GLOBAL.BK	0.05680652
4	PTTEP.BK	0.09083886
5	TISCO.BK	0.30083989
6	AOT.BK	0.01953357
7	KTB.BK	0.31757877

น้ำหนักการลงทุนที่ทำให้ Sharpe Ratio ของ portfolio มากที่สุด

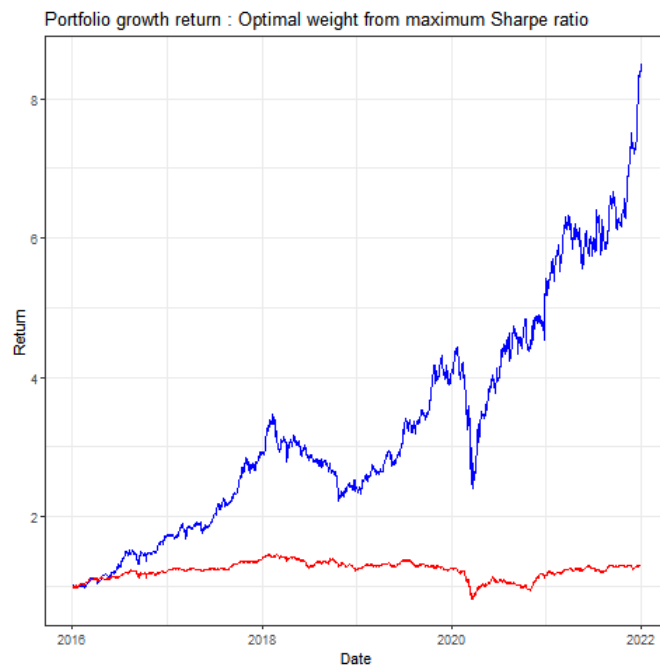
	symbol	weights
1	GPSC.BK	0.388200667
2	JMT.BK	0.419201027
3	GLOBAL.BK	0.051276679
4	PTTEP.BK	0.018826896
5	TISCO.BK	0.089448268
6	AOT.BK	0.006687227
7	KTB.BK	0.026359237

ขั้นตอนที่ 5 : เรียกใช้ฟังก์ชันของขั้นตอนที่ 4 ได้ผลดังนี้

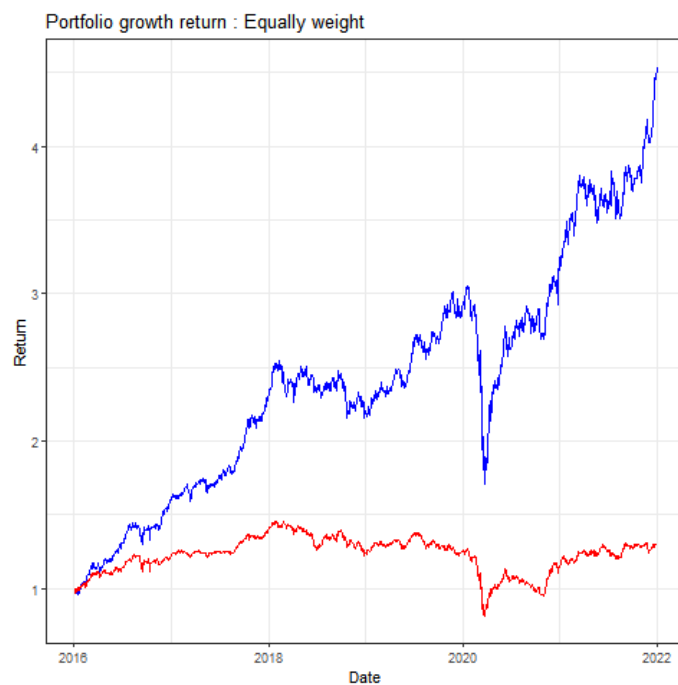
กราฟที่เทียบระหว่าง Portfolio cumulative growth เทียบกับ SET Index Portfolio cumulative growth โดยใช้ข้อมูลที่ทำให้น้ำหนักการลงทุนที่ทำให้ความเสี่ยงของ portfolio น้อยที่สุด



กราฟที่เทียบระหว่าง Portfolio cumulative growth เทียบกับ SET Index Portfolio cumulative growth โดยใช้ข้อมูลที่ทำให้น้ำหนักการลงทุนที่ทำให้ Sharpe Ratio ของ portfolio มากที่สุด



กราฟที่เทียบระหว่าง Portfolio cumulative growth เทียบกับ SET Index Portfolio cumulative growth โดยใช้ข้อมูลที่ทำให้น้ำหนักในการลงทุนของหุ้นทุกตัวใน portfolio เท่ากันทั้งหมด



ผลการดำเนินงาน

1) จัดกลุ่ม 7 กลุ่มด้วยวิธี k-means clustering จัดกลุ่มข้อมูลออกมาได้ทั้งหมด 7 กลุ่ม ดังนี้

กลุ่มที่ 1 : มีหุ้นทั้งหมด 7 ตัว คือ BEM, BTS, CPN, GPSC, KTC, MTC, SAWAD

กลุ่มที่ 2 : มีหุ้นทั้งหมด 3 ตัว คือ EA, JMART, JMT

กลุ่มที่ 3 : มีหุ้นทั้งหมด 6 ตัว คือ ADVANCE, CBG, GLOBAL, HMPRO, INTUCH, KCE

กลุ่มที่ 4 : มีหุ้นทั้งหมด 14 ตัว คือ BANPU, BDMS, BH, CPALL, CPF, EGCO, IRPC, IVL, PTT, PTTEP, PTTGC, SCC, TOP, TU

กลุ่มที่ 5 : มีหุ้นทั้งหมด 3 ตัว คือ DTAC, LH, TISCO

กลุ่มที่ 6 : มีหุ้นทั้งหมด 2 ตัว คือ AOT, MINT

กลุ่มที่ 7 : มีหุ้นทั้งหมด 6 ตัว คือ BBL, BLA, KBANK, KTB, TRUE, TTB

2) เลือกหุ้นที่ดีที่สุดจากทั้ง 7 กลุ่มด้วย Sharpe Ratio และหา optimal portfolio ทั้งสองแบบได้ดังนี้

Portfolio 1 :

ให้อัตราผลตอบแทนของกลุ่มหลักทรัพย์เท่ากับ 20.80%

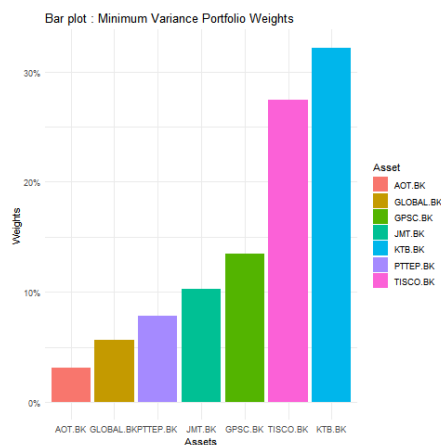
ให้ความเสี่ยงของกลุ่มหลักทรัพย์เท่ากับ 19.38%

ให้ค่า Sharpe Ratio ของกลุ่มหลักทรัพย์เท่ากับ 1.07

และมีน้ำหนักในการลงทุน ดังตารางต่อไปนี้

Stock	Weight
Global Power Synergy (GPSC)	13.47%
JMT Network Services (JMT)	10.30%
Siam Global House (GLOBAL)	5.65%
PTT Exploration and Production Public Company Limited (PTTEP)	7.82%
Tisco Bank (TISCO)	27.44%
Airports of Thailand (AOT)	3.13%
Krung Thai Bank Public Company Limited (KTB)	19.38%

กราฟของน้ำหนักการลงทุนดังภาพ



Portfolio 2 :

ให้อัตราผลตอบแทนของกลุ่มหลักทรัพย์เท่ากับ 40.60%

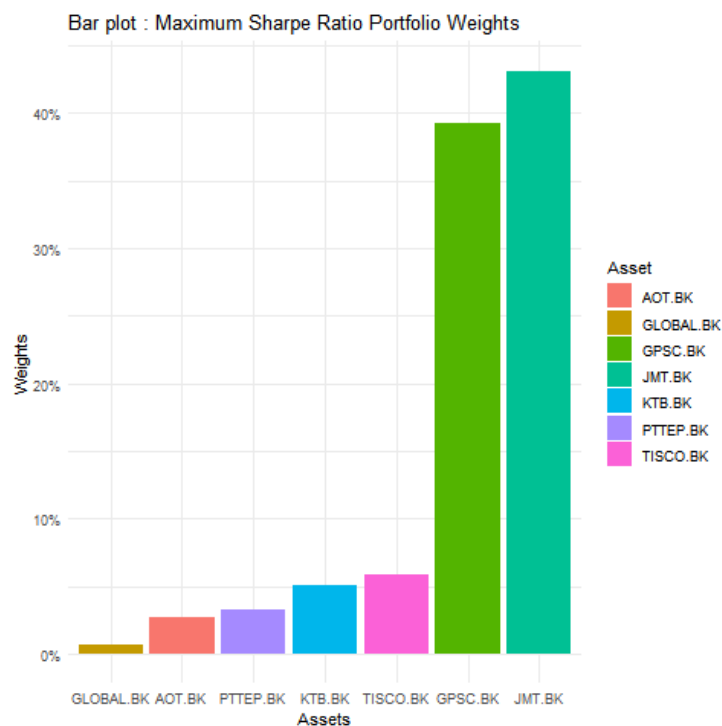
ให้ความเสี่ยงของกลุ่มหลักทรัพย์เท่ากับ 23.60%

ให้ค่า Sharpe Ratio ของกลุ่มหลักทรัพย์เท่ากับ 1.72

และมีน้ำหนักในการลงทุน ดังตารางต่อไปนี้

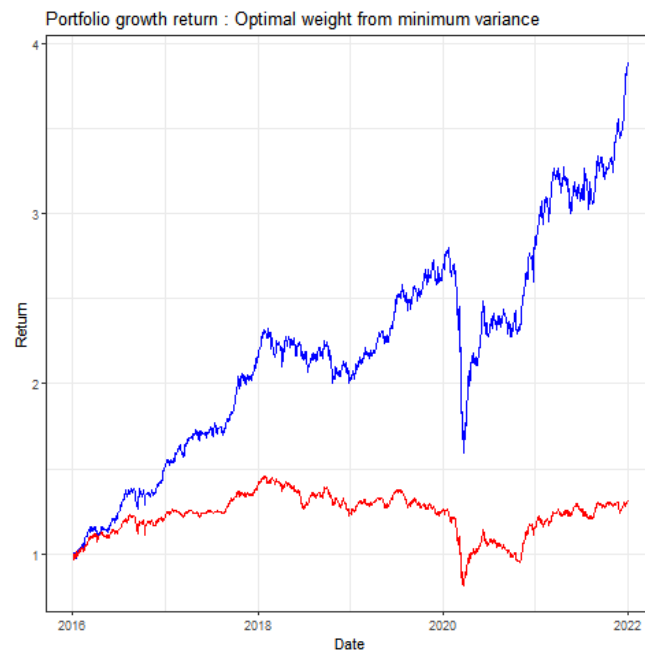
Stock	Weight
Global Power Synergy (GPSC)	39.3%
JMT Network Services (JMT)	43.1%
Siam Global House (GLOBAL)	0.74%
PTT Exploration and Production Public Company Limited (PTTEP)	3.24%
Tisco Bank (TISCO)	5.87%
Airports of Thailand (AOT)	2.74%
Krung Thai Bank Public Company Limited (KTB)	5.04%

กราฟน้ำหนักการลงทุนดังภาพ

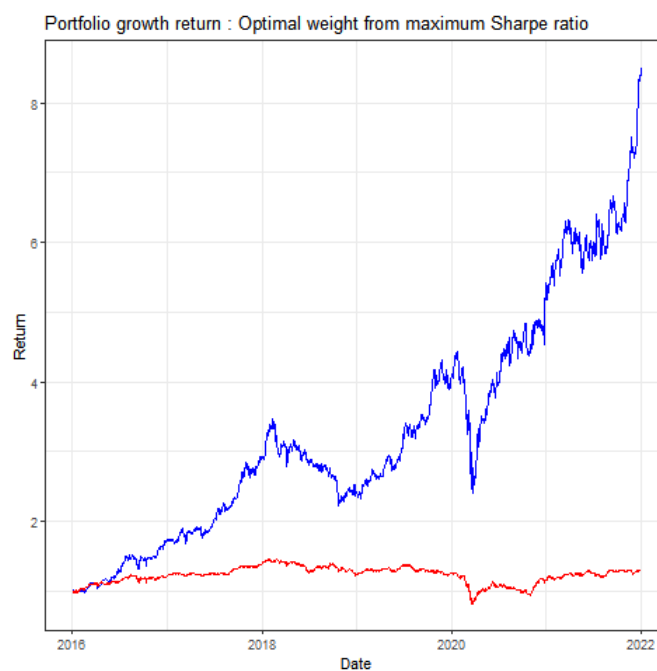


3) กราฟผลจากการเปรียบเทียบอัตราผลตอบแทนสะสม (cumulative return) ของ portfolio ที่มีน้ำหนักที่เหมาะสมตามทฤษฎีพอร์ตการลงทุนสมัยใหม่ กับที่มีน้ำหนักการลงทุนเท่ากัน (Equally weight portfolio) เทียบกับอัตราผลตอบแทนสะสม (cumulative return) ของ SET Index

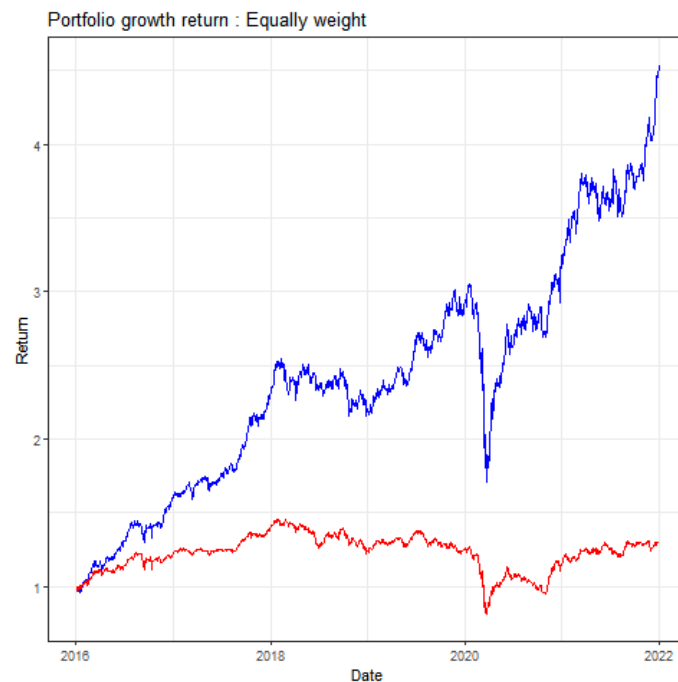
กราฟที่เทียบระหว่าง Portfolio cumulative growth เทียบกับ SET Index Portfolio cumulative growth โดยใช้ข้อมูลที่ทำให้น้ำหนักการลงทุนที่ทำให้ความเสี่ยงของ portfolio น้อยที่สุด



กราฟที่เทียบระหว่าง Portfolio cumulative growth เทียบกับ SET Index Portfolio cumulative growth โดยใช้ข้อมูลที่ทำให้น้ำหนักการลงทุนที่ทำให้ Sharpe Ratio ของ portfolio มากที่สุด



กราฟที่เทียบระหว่าง Portfolio cumulative growth เทียบกับ SET Index Portfolio cumulative growth โดยใช้ข้อมูลที่ทำให้น้ำหนักในการลงทุนของหุ้นทุกตัวใน portfolio เท่ากันทั้งหมด



ดังนั้น จากทั้ง 3 กราฟที่แสดง Portfolio cumulative growth เทียบกับ SET Index Portfolio cumulative growth โดยใช้ข้อมูลที่ทำให้น้ำหนักการลงทุนที่ทำให้ Sharpe Ratio ของ portfolio มากที่สุด ทำให้อัตราผลตอบแทนสะสมได้ดีมากที่สุด

ประโยชน์ที่ได้จากการวิจัย

1. มี Model ที่ทำให้ได้รับหุ้นที่ได้จากการจัดกลุ่มจากเทคนิค k-means clustering และเลือกหุ้นที่ดีที่สุดจากแต่ละกลุ่มด้วย Sharpe Ratio และสามารถนำหุ้นเหล่านี้ไปจัด Portfolio ได้
2. มี Model ที่สามารถหา Optimal weighted portfolio ได้
3. มี Model ที่สามารถ plot line graph ของเพื่อดูการเปรียบเทียบอัตราผลตอบแทนสะสม (cumulative return) ของ portfolio ที่มีน้ำหนักที่เหมาะสมตามทฤษฎีพอร์ตการลงทุนสมัยใหม่ กับที่มีน้ำหนักการลงทุนเท่ากัน (Equally weight portfolio) เทียบกับอัตราผลตอบแทนสะสม (cumulative return) ของ SET Index

เอกสารอ้างอิง

- [1] นภาพรณ จันต๊ะ. (2557). Logistic regression and its application in credit scoring. สืบค้นเมื่อ 25 พฤศจิกายน 2565, จาก http://ethesisarchive.library.tu.ac.th/thesis/2014/TU_2014_5602034224_563_731.pdf?fbclid=IwAR05zjKlcxZ_JJ1SDBYswPIYV_tE0Q4nKZlQW1_LEopMg08QxgVKlum9VXs
- [2] พสธร ฤกษ์พัฒนกิจ, อภิชาติพงศ์สุพัฒน์. (2564). การจัดกลุ่มหลักทรัพย์ลงทุนรายหมวดธุรกิจในตลาดหลักทรัพย์แห่งประเทศไทย. คณะบริหารธุรกิจ มหาวิทยาลัยเกษตรศาสตร์
- [3] Karina Marvin. (2558). Creating Diversified Portfolios Using Cluster Analysis. สืบค้นเมื่อ 25 พฤศจิกายน 2565, จาก https://www.cs.princeton.edu/sites/default/files/uploads/karina_marvin.pdf?fbclid=IwAR3hzUJyT4fswSBrpOQZAp3dq_LThGfhZSBPFCCdGCyHEL1T8vLXp30drdo
- [4] Pakhapoom Sarapat, PhD. (2563). อีกขั้นของ k-means algorithm ที่สามารถแบ่งกลุ่มข้อมูลได้ทุกประเภท.. สืบค้นเมื่อ 21 พฤศจิกายน 2565, จาก https://bigdata.go.th/big-data-101/k-means-algorithm-for-clustering-large-data-sets-with-categoricalvalues/?fbclid=IwAR1ReIWqTslngZ-gycAwkn9NjCG8x9ivWgxTF2JngXpaB2D7TsClGWypZ_Y
- [5] University of Cincinnati. K-means Cluster Analysis. สืบค้นเมื่อ 21 พฤศจิกายน 2565, จาก https://uc-r.github.io/kmeans_clustering?fbclid=IwAR0-FYXg1UC46MQwq5X8OXhABrZiLasvNQ0Phv5x4Wb1aWd7T9G2zbPgZqA