Journey

* Data set ที่ได้รับมามีรูปแบบ table ที่ ไม่เหมาะสำหรับการ เปลี่ยนเป็น data frame จึงต้องทำการ Trim และ Rename column ใหม่ทั้งหมด
* ตรวจสอบ data set info(), nunique() , describe() , isna().sum() พบว่า มี ข้อมูลอยู่ 3 แถวที่ ไม่มีข้อมูล sector จึงเลือกใช้คำสั่ง dropna() เนื่องจาก ไม่มีนัยสำคัญ
* เลือก slice column ที่ต้องการนำมา วิเคราะห์ case Q2/65 ,Q2/64 และ 6m65 , 6m64 แยก ออกเป็น 2 cases พบปัญหาคือ value ใน column บางแถว มีสถานะเป็น object จริงใช้คำสั่ง pd.to\_numeric() เพื่อแปลงค่า เพื่อที่จะใช้ คำนวณได้ต่อไป
* พบปัญหา **ValueError**: Unable to parse string "(แก้ไข)" at position 449 ทำการแก้ไขโดย เปลี่ยน value ช่องนั้นให้กลายเป็น )0.0 ด้วย df.replace()
* เนื่องจากการวิเคราะห์จะให้น้ำหนักกับปัจจุบันมากกว่าจึงนำ เลือกcolumn ที่มีข้อมูลเกี่ยวข้องกับ ไตรมาส2 มาวิเคราะห์ต่อ โดยแบ่งออกเป็น eps growth/Q265 เป็น บวก และ eps growth/Q265 เป็นลบ
* นำกลุ่ม eps growth/Q265 เป็น บวก มาจับกลุ่ม group by sector และเรียงลำดับ แล้วนับจำนวน
* Plot กราฟ จำนวนหุ้น แบ่งตาม sector และ ลองทดสอบการ Plot earning/Q265 เทียบ earning/Q264 ของหุ้นแต่หละตัวแยกตาม sector ของตัวบริษัท
* นำ for loop มาประยุกต์ใช้ เพื่อให้ได้ กราฟของทุก sector พบปัญหา ไม่สามารถทำ subplot ได้สวยงาม และลองใช้ sns.catplot() แต่ไม่สามารถใช้ได้เนื่องจาก แต่หละ sector มี บริษัทที่หลากหลาย ไม่เท่ากัน
* เริ่มวิเคราะห์ earning growth Q2/65/64 VS 6month/65/64 โดย slice column ที่ต้องการ และนำมาต่อกันโดย pd.concat() และนำ Df ที่ได้มา filter ด้วย q2\_65epsg > 0.0 & 6m\_65epsg >= mean
* Plot graph จาก DF ใหม่ที่ได้ โดยใช้ for loop เพื่อให้ได้ กราฟของทุก sector
* กลับมาดู Df ที่ eps growth/Q265 เป็นลบ สร้าง Df ใหม่ ที่มีค่า q2\_65eps ลบด้วย q2\_64eps แล้ว สร้างDf ใหม่ที่ ประกอบด้วย sector company และ q2\_65eps ลบด้วย q2\_64eps
* นำ Df ใหม่ที่ได้ มา filter หา q2\_65eps ลบด้วย q2\_64eps เป็น บวก เพื่อหาบริษัทที่ มีความเป็นไปได้ที่จะturn around กลับมาจาก ผลประกอบการเป็นลบ
* นำ Df ที่ได้มา Visualize ด้วยคำสั่ง sns.catplot() แต่พบปัญหาคือ มีบริษัทที่มี q2\_65eps เป็นลบ และ q2\_64eps เป็นลบ จึงต้องแก้ไขอีกครั้ง
* นำ Df ที่ทำงานอยู่ มารวมกับ Df ที่ประกอบด้วย sector company q2\_65earn q2\_sub ด้วยคำสั่ง pd.merge() และเลือก filter เฉพาะ q2\_65earn เป็น บวกเท่านั้นและทำ Visualizeอีกครั้ง
* Ranked companies by sector with earning per share growth and compared between 6m/ปี65และ64
* สร้าง Df ที่ประกอบด้วย q2\_65earn + q2\_64earn = earnsum และ 6m\_65earn + 6m\_64earn = 6m\_earnsum เพื่อที่จะนำมา วิเคราะห์หา growth
* นำ Df มา filter เฉพาะ earnsum และ 6m\_earnsum ที่อยู่เหนือค่า mean
* นำ Df ที่ได้มา transform ค่าด้วยวิธี standardization เนื่องจาก range ของข้อมูลค่อนข้างกว้าง
* สร้าง column ใหม่ earnsum หารด้วย 6m\_earnsum คูณด้วย 100 แปลงค่าเป็น percentage
* Concat ข้อมูลใหม่ที่ได้กลับไปรวมกับ Df sector company และ filter หา case >= 100.00
* นำ Df ที่ได้มา Visualize แบ่งตาม sector อีกครั้ง