บทที่ 3
วิธีดำเนินโครงงาน

ในการพัฒนาโครงการต้องมีการวิเคราะห์และออกแบบข้อมูลข้อมูลที่ได้มาให้สอดคล้องกับความต้องการของผู้ใช้เพื่อให้เกิดประโยชน์สูงสุด และทำให้โครงการสำเร็จลุล่วง ผู้วิเคราะห์จึงได้วิเคราะห์ และออกแบบเครื่องข้อมูลใหม่ โดยใช้เครื่องมือในการออกแบบ Tableau และกระบวนการในการพัฒนาฐานข้อมูล CRISP-DM สร้างแบบฟอร์มข้อมูล และปรับโครงสร้างข้อมูลใหม่ ดังนี้

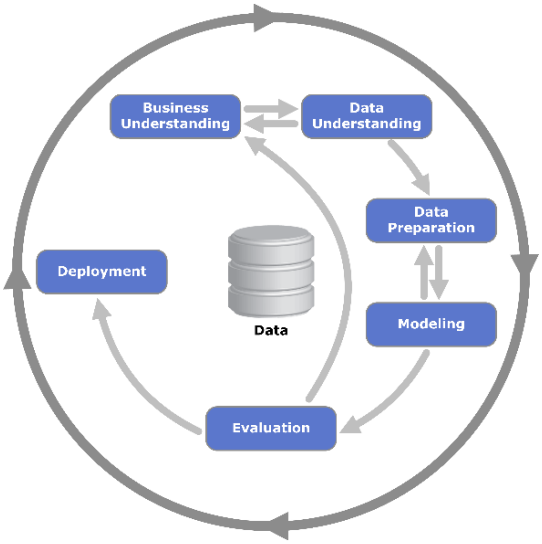
3.1 กระบวนการ CRISP-DM

3.2 การออกแบบและพัฒนาเว็บไซต์

3.3 สรุป

3.1 กระบวนการ CRISP-DM

กระบวนการ CRISP-DM นี้จะประกอบด้วย 6 ขั้นตอน แต่ละขั้นตอนในรูปจะเป็นขั้นตอนที่ต่อเนื่องกันนั่นคือขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้าซึ่งแสดงด้วยลูกศรที่เชื่อมระหว่างกล่องสี่เหลี่ยมแต่ละกล่อง ตัวอย่างเช่นเมื่อได้ผลลัพธ์จากขั้นตอนการเตรียมข้อมูล (Data Preparation) แล้วจะนำไปสร้างโมเดลจำแนกประเภทข้อมูลในขั้น Modeling และหลังจากนั้นอาจจะย้อนกลับมาเปลี่ยนแปลงข้อมูลให้ถูกต้องมากขึ้นเพื่อหวังว่าจะโมเดลที่ให้ความถูกต้องมากขึ้นก็ได้ เป็นต้น



ภาพที่ 3.1 แสดงกระบวนการ CRISP-DM

ที่มา : dataminingtrend.com (ม.ป.ป)

ในกระบวนการนี้ประกอบด้วย 6 ขั้นตอน

1) Business Understanding เป็นขั้นตอนที่เน้นการเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูลทาง Data Mining พร้อมทั้งวางแผนในการดำเนินการคร่าว ๆ

2) Data Understanding เป็นขั้นตอนการเก็บรวบรวม และตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้ เพื่อตรวจสอบความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

3) Data Preparation เป็นขั้นตอนที่ทำการแปลงข้อมูลที่เก็บรวบรวมมาให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลให้อยู่ในช่วงเดียวกัน หรือการเติมข้อมูลที่ขาดหายไป

4) Modeling เป็นขั้นตอนที่ผู้วิเคราะห์ทำการวิเคราะห์ข้อมูลที่ได้จากขั้นตอนก่อนหน้าด้วยเทคนิคแบบจำลองอนุกรมเวลา (Time Series)

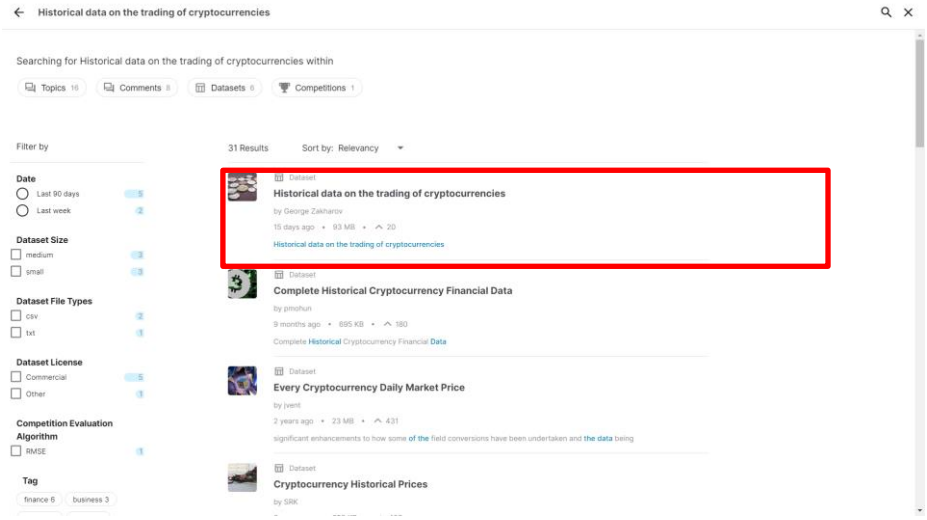
5) Evaluation เป็นขั้นตอนที่ผู้วิเคราะห์จะทำการตรวจวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ต้องไว้แต่แรกหรือไม่ แล้วจึงนำผลลัพธ์ที่ได้ไปใช้งานต่อ

6) Deployment มีการนำความรู้ที่ได้จากการได้ผลลัพธ์ด้วยเทคนิค Data Mining ไปใช้ประโยชน์ต่อไป

3.1.1 ความเข้าใจในธุรกิจ (Business Understanding) ผู้วิเคราะห์ได้ทำความเข้าใจกระบวนการทางข้อมูลและระบบวิเคราะห์ข้อมูลกับปัญหาของข้อมูลราคาหุ้นในตลาดหลักทรัพย์ เพื่อที่จะนำผลการวิเคราะห์ข้อมูลไปใช้ให้เกิดประโยชน์และวางแผนในการดำเนินการโดยการวิเคราะห์ ข้อมูลรูปแบบความสัมพันธ์ที่อยู่ในข้อมูลของข้อมูลราคาหุ้นคริปโตเคอร์เรนซี

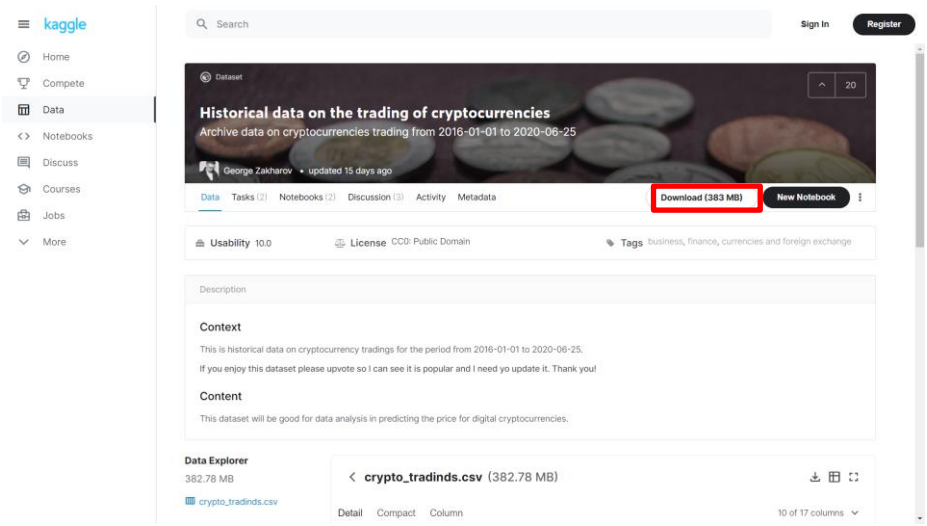
3.1.2 การทำความเข้าใจข้อมูล (Data Understanding) ผู้วิเคราะห์ข้อมูลเริ่มทำการเก็บรวบรวมข้อมูล ตรวจสอบข้อมูลที่ได้ดูความถูกต้องและความน่าเชื่อถือของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลราคาหุ้นคริปโตเคอร์เรนซีบางส่วนมาใช้ในการวิเคราะห์ ชุดข้อมูล https://www.kaggle.com ซึ่งผู้ใช้ชื่อว่า gezakharov ได้ทำการรวบรวมและเผยแพร่ผู้วิเคราะห์จึงได้ดำเนินการตามขั้นตอนดังนี้

1) ผู้วิเคราะห์ข้อมูลรวบรวมข้อมูลจากเว็บไซต์ https://www.kaggle.com โดยได้ทำการเลือกชุดข้อมูล Historical data on the trading of cryptocurrencies



ภาพที่ 3.2 แสดงขั้นตอนการหา Big Data จากเว็บไซต์

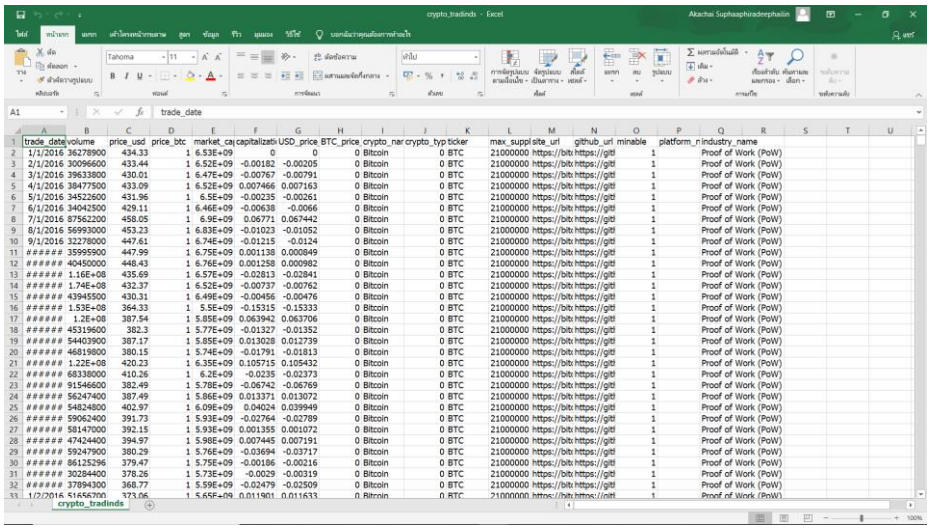
2) ผู้วิเคราะห์ทำการดาวน์โหลดเพื่อเก็บรวบรวมข้อมูล



ภาพที่ 3.3 แสดงขั้นตอนการดาวน์โหลด

3) ตรวจสอบความถูกต้อง และความน่าเชื่อถือของข้อมูล ซึ่ง 1 ไฟล์ จะประกอบด้วย 17 แอตทริบิวท์ คือ วันที่ , ปริมาณซื้อขาย , ราคาต่อ USD , ราคาต่อ BTC , มูลค่าหลักทรัพย์ตามราคาตลาด , มูลค่าหลักทรัพย์ที่เปลี่ยนแปลงก่อนหน้าหนึ่งวัน , มูลค่า USD ที่เปลี่ยนแปลงก่อนหน้าหนึ่งวัน , มูลค่า BTC ที่เปลี่ยนแปลงก่อนหน้าหนึ่งวัน , ชื่อหุ้น ,

ประเภทของหุ้น , ชื่อย่อของหุ้น , จำนวนหุ้นทั้งหมดที่จะเกิดขึ้น , เว็บไซต์ของหุ้น , เว็บไซต์ Github ของหุ้น , minable , ชื่อแพลตฟอร์ม , ชื่ออุตสาหกรรม



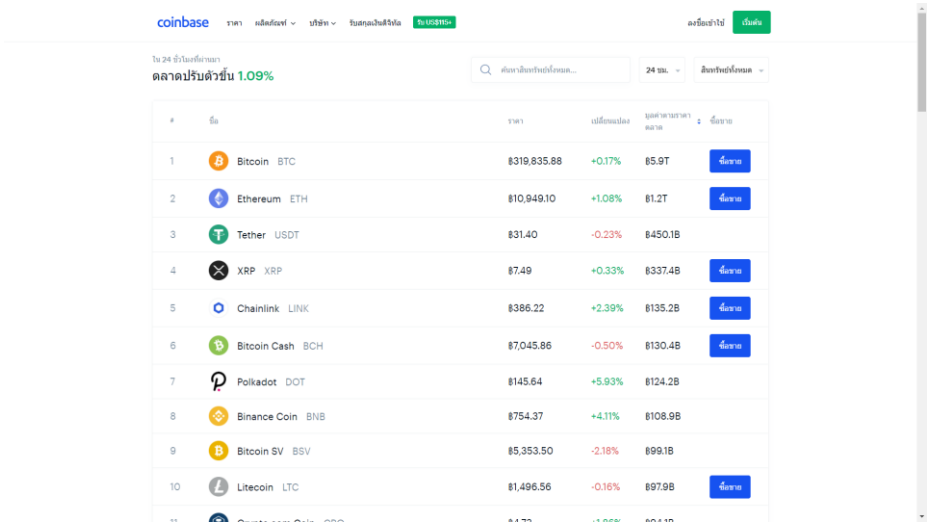
trade_date	volume	price_usd	price_btc	market_ca	capitalizati	USD_price	BTC_price	crypto_nar	crypto_typ	ticker	max_suppl	site_url	github_url	minable	platform_n	industry_name
1/1/2016	36278900	434.33	1	6.53E+09	0	0	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
2/1/2016	30096600	433.44	1	6.52E+09	-0.00182	-0.00205	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
3/1/2016	39633800	430.01	1	6.47E+09	-0.00767	-0.00791	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
4/1/2016	38477500	433.09	1	6.52E+09	0.007466	0.007163	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
5/1/2016	34522600	431.96	1	6.5E+09	-0.00235	-0.00261	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
6/1/2016	34042500	429.11	1	6.46E+09	-0.00638	-0.0066	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
7/1/2016	87562200	458.05	1	6.9E+09	0.06771	0.067442	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
8/1/2016	56993000	453.23	1	6.83E+09	-0.01023	-0.01052	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
9/1/2016	32278000	447.61	1	6.74E+09	-0.01215	-0.0124	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	35995900	447.99	1	6.75E+09	0.001138	0.000849	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	40450000	448.43	1	6.76E+09	0.001258	0.000982	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	1.16E+08	435.69	1	6.57E+09	-0.02813	-0.02841	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	1.74E+08	432.37	1	6.52E+09	-0.00737	-0.00762	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	43945500	430.31	1	6.49E+09	-0.00456	-0.00476	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	1.53E+08	364.33	1	5.5E+09	-0.15315	-0.15333	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	1.2E+08	387.54	1	5.85E+09	0.063942	0.063706	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	45319600	382.3	1	5.77E+09	-0.01327	-0.01352	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	54403900	387.17	1	5.85E+09	0.013028	0.012739	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	46819800	380.15	1	5.74E+09	-0.01791	-0.01813	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	1.22E+08	420.23	1	6.35E+09	0.105715	0.105432	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	68338000	410.26	1	6.2E+09	-0.0235	-0.02373	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	91546600	382.49	1	5.78E+09	-0.06742	-0.06769	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	56247400	387.49	1	5.86E+09	0.013371	0.013072	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	54824800	402.97	1	6.09E+09	0.04024	0.039949	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	59062400	391.73	1	5.93E+09	-0.02764	-0.02789	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	58147000	392.15	1	5.93E+09	0.001355	0.001072	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	47424400	394.97	1	5.98E+09	0.007445	0.007191	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	59247900	380.29	1	5.76E+09	-0.03694	-0.03717	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	86125296	379.47	1	5.75E+09	-0.00186	-0.00216	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	30284400	378.26	1	5.73E+09	-0.0029	-0.00319	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)
######	37894300	368.77	1	5.59E+09	-0.02479	-0.02509	0	Bitcoin	0	BTC	21000000	https://bitc	https://gitl	1		Proof of Work (PoW)

ภาพที่ 3.4 ข้อมูลบันทึกหุ้นคริปโตเคอร์เรนซี

ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมา ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมาก ระบบการรับข้อมูลป้อนเข้าสู่ระบบในปัจจุบันจะลดการคีย์ข้อมูลจากคนให้น้อยที่สุด แต่จะใช้วิธีการสแกน การติ๊กเลือก เพื่อลดความผิดพลาดให้น้อยที่สุด เพราะขั้นตอนใช้เวลามากการลดข้อผิดพลาดของข้อมูลได้มากเท่าใด ก็จะยิ่งมีประสิทธิภาพมากขึ้นเท่านั้น

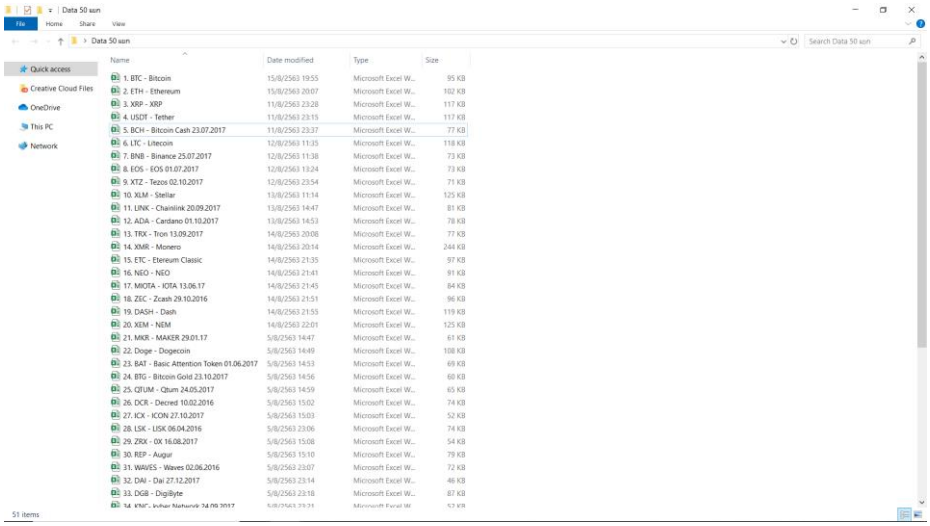
3.1.3 การเตรียมข้อมูล (Data Preparation) ผู้วิเคราะห์ทำการเปลี่ยนข้อมูลที่เก็บรวบรวมให้เป็นรูปแบบเดียวกัน และกำจัดข้อมูลเสียออก เพื่อเตรียมสำหรับการนำข้อมูลไปวิเคราะห์ในขั้นต่อไป โดยใช้กระบวนการ Data Cleaning โดยขั้นตอนการแปลงข้อมูลที่รวบรวมมาให้มีประสิทธิภาพ และอยู่ในรูปแบบที่พร้อมสำหรับการนำไปวิเคราะห์ในขั้นต่อไปได้นั้นเป็นขั้นตอนที่ใช้ระยะเวลานาน

1) ทำการเลือกหุ้นคริปโตเคอร์เรนซี ที่มีข้อมูลตั้งแต่วันที่ 1 มกราคม 2561 ถึง 31 ธันวาคม 2562โดยอ้างอิงอันดับหุ้นคริปโตเคอร์เรนซีที่ผู้คนให้ความสนใจมากที่สุดจาก https://www.coinbase.com/price ณ วันที่ 5 พฤษภาคม 2563 ประกอบด้วย



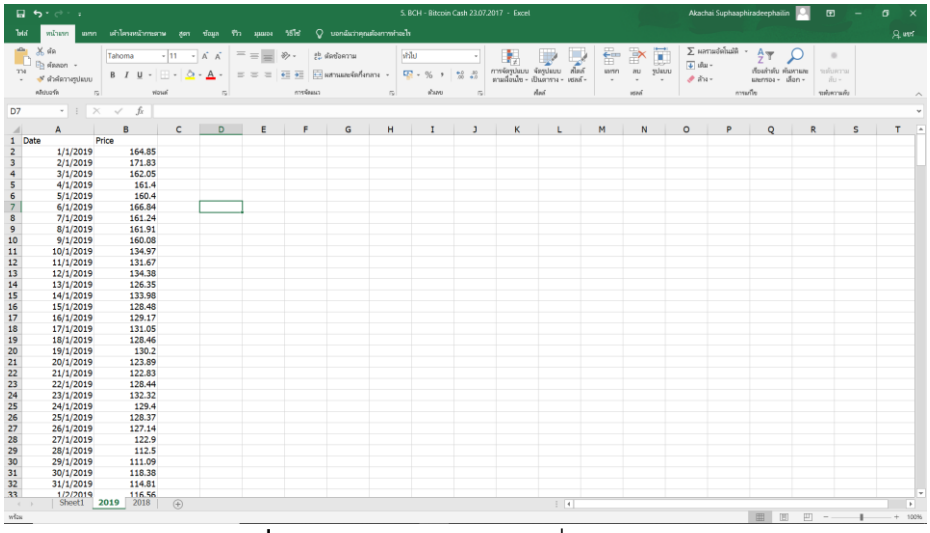
ภาพที่ 3.5 ภาพอันดับหุ้นคริปโตเคอร์เรนซี

2) จากนั้นผู้วิเคราะห์จึงแยกหุ้นคริปโตเคอร์เรนซีแต่ละรายการออกจากกัน



ภาพที่ 3.6 ไฟล์ข้อมูลหุ้นคริปโตเคอร์เรนซี

3) ผู้วิเคราะห์ได้ทำการคัดเลือก แอตทริบิวท์ที่มีความเหมาะสมต่อการพยากรณ์มากที่สุด จึงเหลือเพียง 2 แอตทริบิวท์ คือ วันที่ และราคาต่อ USD และแยกออกเป็นปี 2561 และ 2562

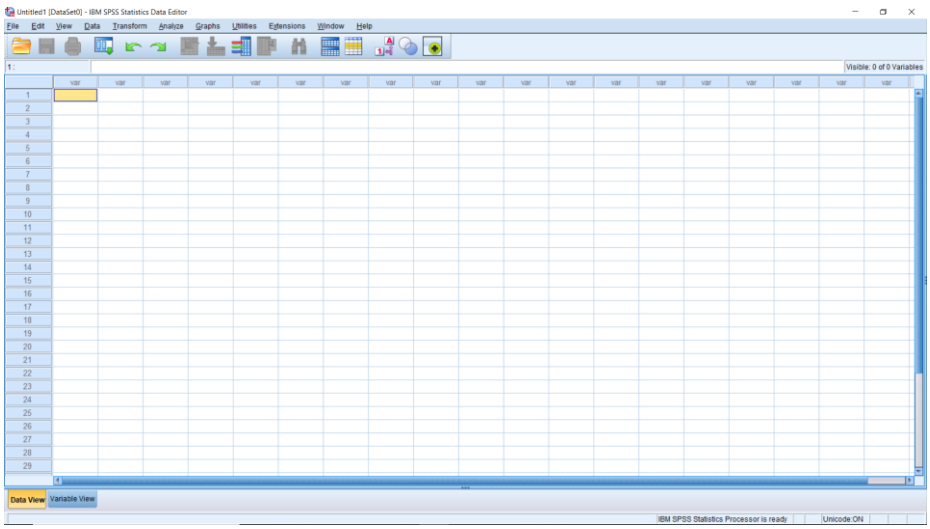


Date	Price
1/1/2019	164.85
2/1/2019	171.83
3/1/2019	162.05
4/1/2019	161.4
5/1/2019	160.4
6/1/2019	166.84
7/1/2019	161.24
8/1/2019	161.91
9/1/2019	160.08
10/1/2019	134.97
11/1/2019	131.67
12/1/2019	134.38
13/1/2019	126.35
14/1/2019	133.98
15/1/2019	128.48
16/1/2019	129.17
17/1/2019	131.05
18/1/2019	128.46
19/1/2019	130.2
20/1/2019	123.89
21/1/2019	122.83
22/1/2019	128.44
23/1/2019	132.32
24/1/2019	129.4
25/1/2019	128.37
26/1/2019	127.14
27/1/2019	122.9
28/1/2019	112.5
29/1/2019	111.09
30/1/2019	118.38
31/1/2019	114.81
1/2/2019	116.56

ภาพที่ 3.7 ข้อมูลแอตทริบิวท์ที่นำไปวิเคราะห์

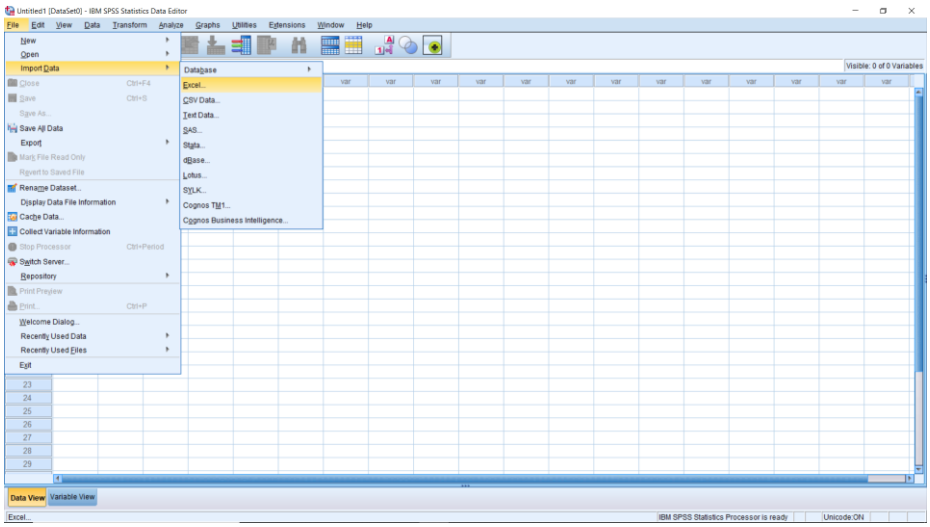
4) หลังจากการเตรียมข้อมูลให้อยู่ในมาตรฐานเดียวกันแล้ว ผู้วิเคราะห์จึงได้เริ่มทำการวิเคราะห์ข้อมูลโดยใช้โปรแกรม IBM SPSS ดำเนินกระบวนการ เพื่อเป็นควบคุมคุณภาพของข้อมูลเพื่อที่จะสามารถนำไปวิเคราะห์และสร้างออกมาเป็นรายงานภาพ Visualization โดยมีขั้นตอนการดำเนินการดังนี้

4.1) โหลดข้อมูลที่ได้เข้าโปรแกรม IBM SPSS เพื่อเตรียมการพยากรณ์



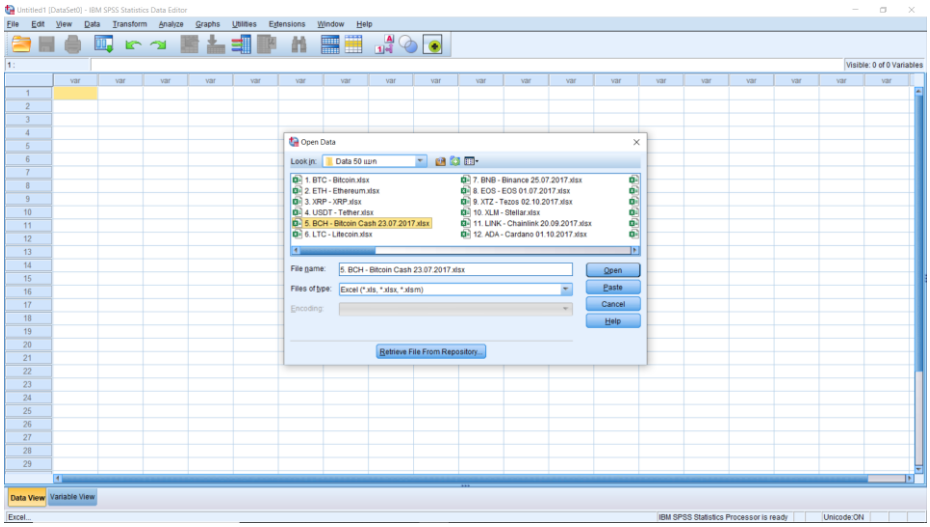
ภาพที่ 3.8 หน้าต่างพื้นที่ทำงานของ IBM SPSS

4.1.1) เลือกคำสั่ง File –> Import Data –> Excel



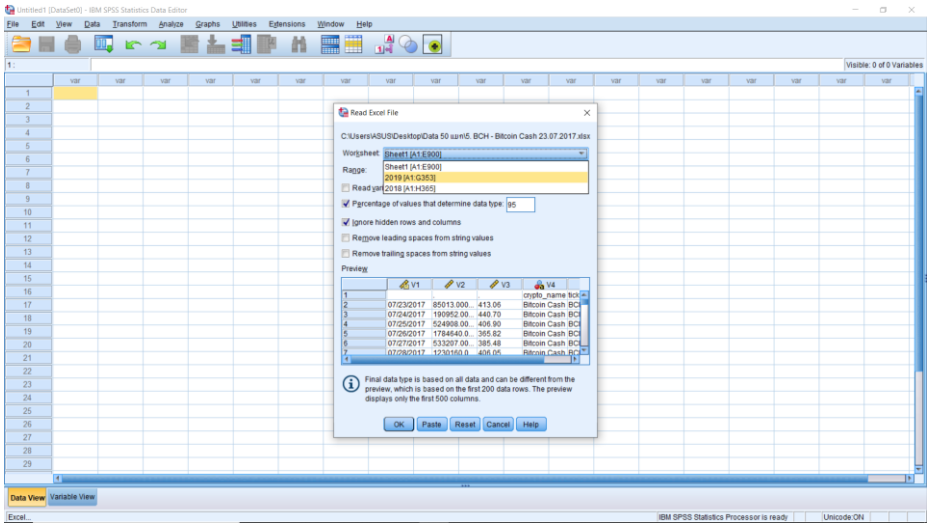
ภาพที่ 3.9 วิธีการ Import ข้อมูล (1)

4.1.2) เลือกชุดข้อมูลที่จะนำคำนวณ



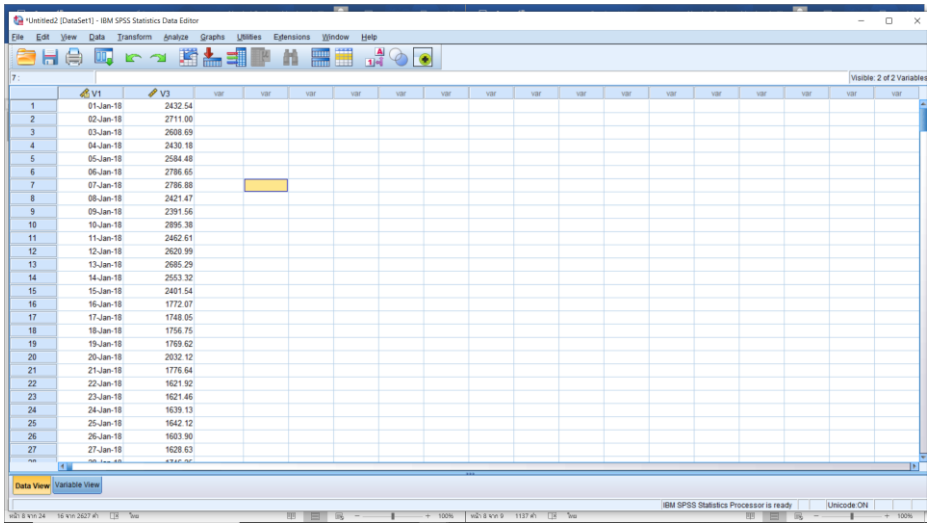
ภาพที่ 3.10 วิธีการ Import ข้อมูล (2)

4.1.3) เนื่องจากผู้วิเคราะห์มีการแบ่งข้อมูลไว้ในไฟล์ Excel ออกเป็น 2 ปี จึงต้องเลือกไฟล์ย่อยที่จะนำมาคำนวณอีกครั้ง

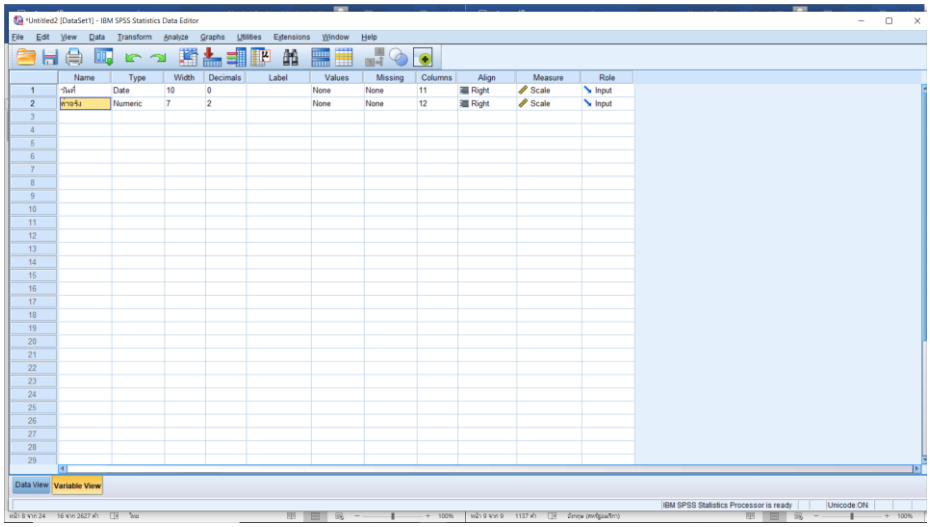


ภาพที่ 3.11 วิธีการ Import ข้อมูล (3)

4.2) ผู้วิเคราะห์ทำการเปลี่ยนชื่อหัวตาราง เพื่อให้ง่ายต่อการเรียกใช้คอลัมน์นั้น ๆ



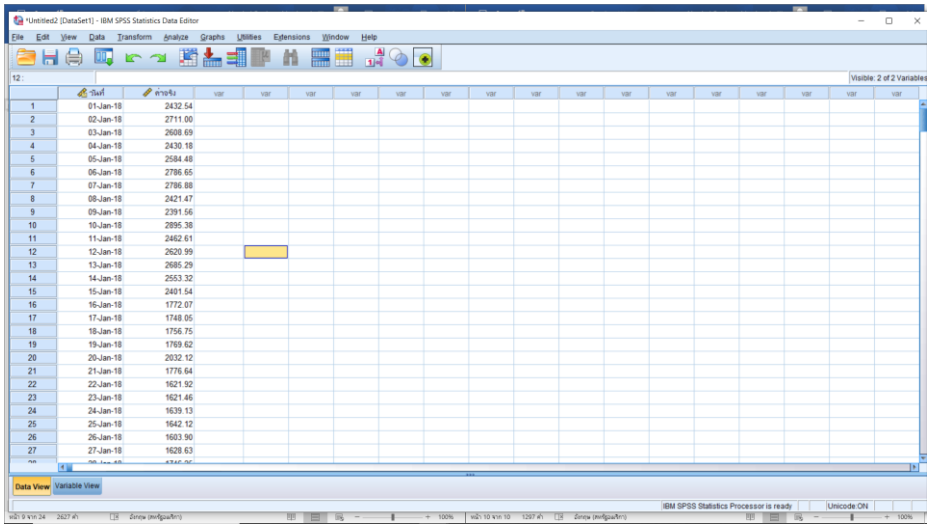
ภาพที่ 3.12 การเปลี่ยนชื่อหัวตาราง (1)



ภาพที่ 3.13 การเปลี่ยนชื่อหัวตาราง (2)

3.1.4 การสร้างแบบจำลอง (Modeling) ผู้วิเคราะห์ข้อมูลวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้า ไมน์นิ่ง Classification เพื่อใช้พยากรณ์ราคาหุ้นคริปโตเคอร์เรนซี โดยใช้การพยากรณ์ในรูปแบบของสมการคณิตศาสตร์ โดยการนำข้อมูลในอดีตมาสร้างรูปแบบของการพยากรณ์ในอนาคต ซึ่งผู้วิเคราะห์ได้เลือกใช้ เทคนิครูปแบบอนุกรมเวลา (Time Series Models) ซึ่งได้แก่ วิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล และแบบจำลองอารีมา โดยใช้โปรแกรมที่ใช้ทำเหมืองข้อมูล ด้วยชุดข้อมูลที่คัดเลือกโดยมีขั้นตอนการดำเนินการ ดังนี้

ในการพยากรณ์โดยใช้รูปแบบอนุกรมเวลาจะเป็นการใช้ข้อมูลในอดีตเพื่อเป็นการพยากรณ์ข้อมูลในอนาคตเท่านั้น โดยจะไม่นำตัวแปรอื่นมาพิจารณา โดยข้อมูลที่นำมาพยากรณ์นั้นต้องมีรูปแบบที่เปลี่ยนไปอย่างมีทิศทาง (Trend)



ภาพที่ 3.14 ตารางข้อมูลตัวอย่าง

จากภาพที่ 3.14ประกอบด้วย 2 แอตทริบิวต์ คือ

1) วันที่ เก็บข้อมูลช่วงเวลาในแต่ละวันที่เป็นระยะเวลา 1 ปี

2) ค่าจริง เก็บข้อมูลราคาหุ้นคริปโตเคอร์เรนซี ในแต่ละวันเป็นระยะเวลา 11 เดือน

การสร้างโมเดล Time Series จะทำการเลือกแอตทริบิวต์ข้อมูลเชิงปริมาณที่จัดเก็บไว้ มาคำนวณหาค่าความเป็นไปได้เชิงพยากรณ์ โดยการคำนวณจากสมการทางคณิตศาสตร์ ดังนี้

ขั้นตอนที่ 1 ผู้วิเคราะห์ทำการคำนวณหาค่าพยากรณ์

วิธีที่ 1 วิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (EMA) เป็นวิธีการสร้างสมการพยากรณ์โดยการหาค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักโดยให้น้ำ หนักของข้อมูลในปัจจุบันมากที่สุด และน้ำหนักจะลดหลั่นกันไปแบบเอ็กซ์โปเนนชียลสำหรับค่าของข้อมูลที่ห่างไกลออกไป โดยมีค่าสัมประสิทธิ์ปรับให้เรียบ α มีค่าอยู่ระหว่าง 0 ถึง 1 วิธีนี้เหมาะสำหรับข้อมูลที่ไม่มีแนวโน้มและฤดูกาลมีสมการพยากรณ์หนึ่งช่วงเวลาล่วงหน้าที่เวลา t ดังนี้

$$\hat{Y}_t(1) = \alpha Y_t + (1-\alpha)\hat{Y}_{t-1}(1)$$

$$= \hat{Y}_{t-1}(1) + \alpha(Y_t - \hat{Y}_{t-1}(1))$$

$$\text{จะได้ } = \hat{Y}_{t-1}(1) + \alpha e_t$$

เมื่อ Y_t คือ ข้อมูลจริง ณ เวลา t

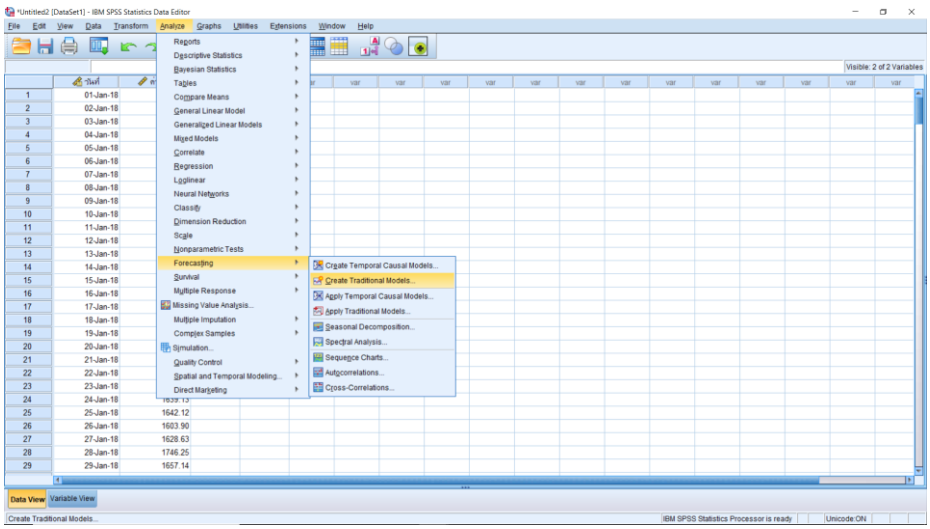
$Y_t(1)$ คือ ค่าพยากรณ์ณ เวลา t+1

$Y_{t-1}(1)$ คือ ค่าพยากรณ์ที่ผ่านมาล่าสุด

e_t คือ ผลต่างระหว่างค่าจริงกับค่าพยากรณ์

ผู้วิเคราะห์ทำการเตรียมข้อมูลทั้งหมด 11 เดือน และทำการตั้งข้อมูลให้เป็นวันที่ 1–334 เพื่อทำการพยากรณ์หาข้อมูลของเดือนที่ 12 สำหรับการนำมาเปรียบเทียบกับข้อมูลจริง

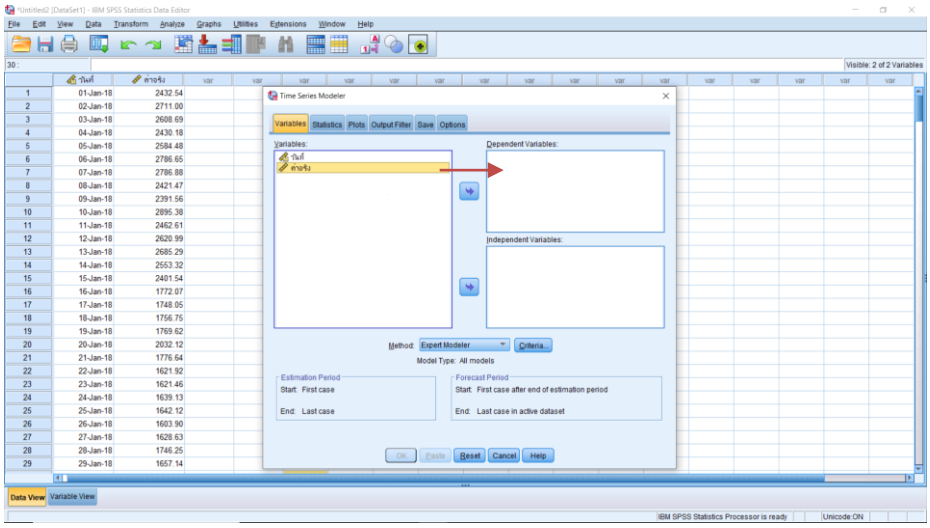
1) เปิดหน้าต่างสำหรับการตั้งค่าการพยากรณ์ เลือกคำสั่ง Analyze -> Forecasting-> Create Traditional Models



ภาพที่ 3.15 วิธีเปิดหน้าต่างสำหรับการตั้งค่าการพยากรณ์

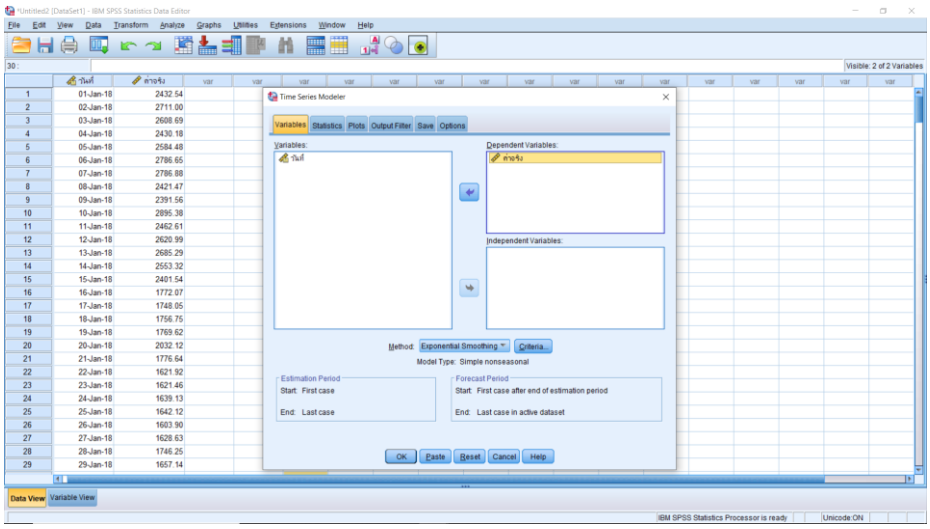
2) การตั้งค่าการพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (EMA)

2.1) เลือกค่าจริงใส่ในช่อง Dependent Variables



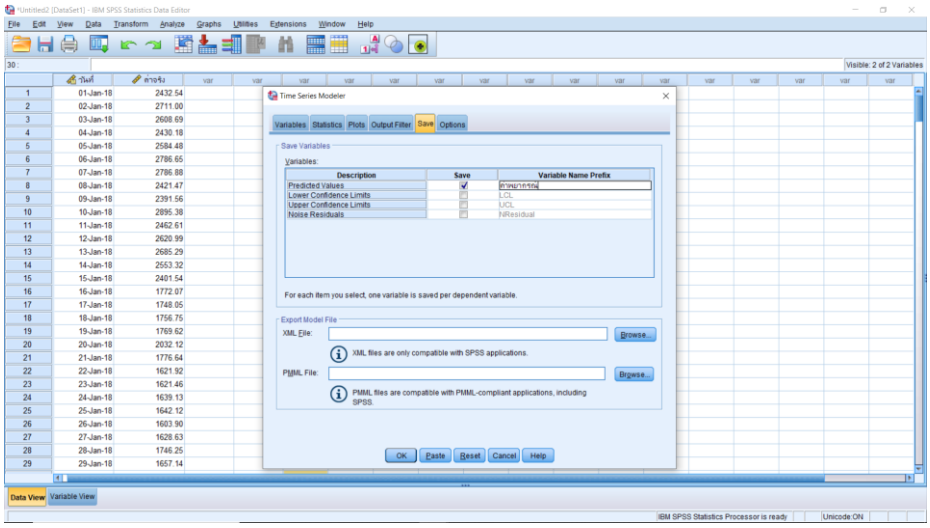
ภาพที่ 3.16 การตั้งค่าการพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (1)

2.2) เลือก Method เป็น Exponential Smoothing



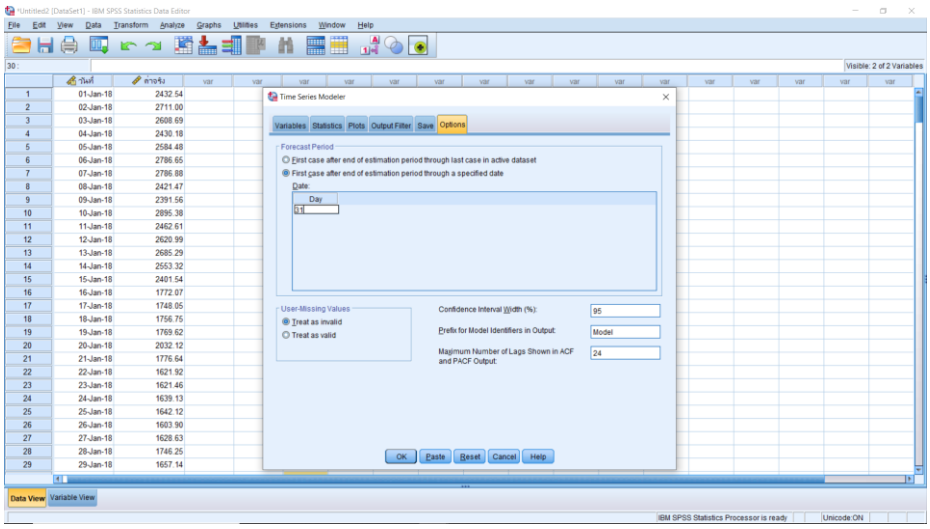
ภาพที่ 3.17 การตั้งค่าการพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (2)

2.3) ทำการเลือกบันทึก Predicted Values และตั้งชื่อ



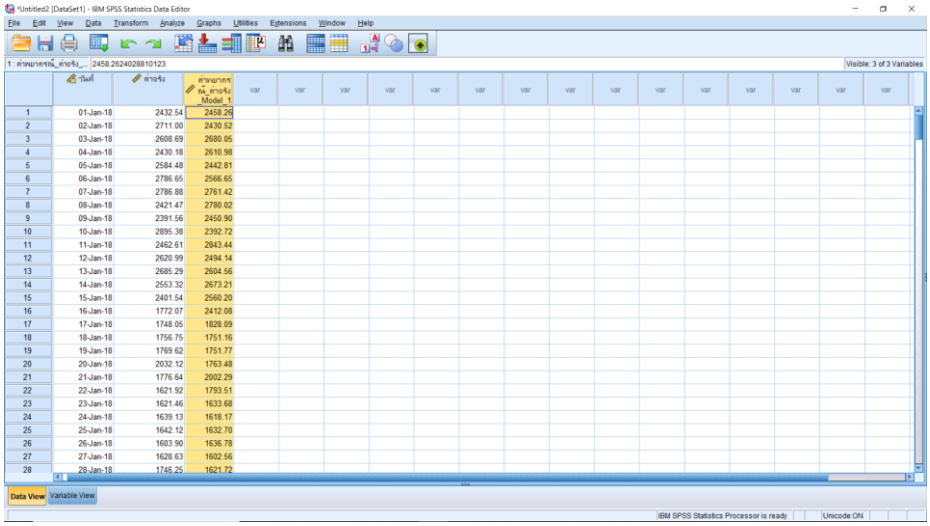
ภาพที่ 3.18 การตั้งค่าการพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (3)

2.4) ใส่จำนวนวันเท่ากับ 31 วันเนื่องจากเดือนธันวาคมมี 31 วัน

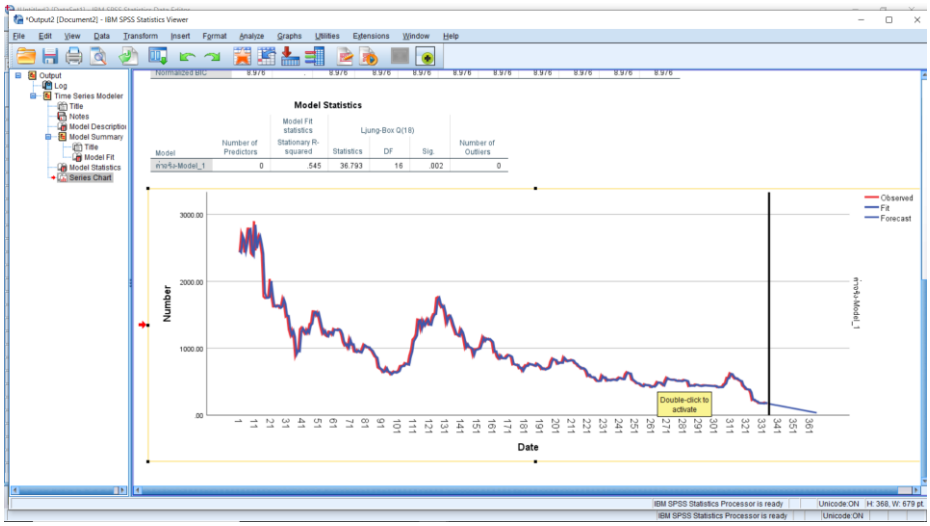


ภาพที่ 3.19 การตั้งค่าการพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (4)

2.5) ค่าพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียลล่วงหน้า 31 วัน



ภาพที่ 3.20 แสดงค่าพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียลล่วงหน้า 31 วัน



ภาพที่ 3.21 แสดงค่าพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียลล่วงหน้า 31 วัน

จากภาพที่ 3.20 เส้นสีแดง หมายถึงค่าจริง

เส้นสีน้ำเงิน หมายถึงค่าพยากรณ์

เส้นสีฟ้า หมายถึงค่าพยากรณ์ล่วงหน้าที่โปรแกรม IBM SPSS คำนวณ

วิธีที่ 2 อารีมา (Arima)

เป็นวิธีที่ได้รับความนิยมอย่างมาก เนื่องจากวิธีการพยากรณ์นี้สามารถใช้ได้กับข้อมูลที่มีลักษณะการเคลื่อนไหวทุกประเภท และให้ผลการพยากรณ์ที่ค่อนข้างแม่นยำกว่าวิธีการพยากรณ์อื่น ๆ สำหรับการพยากรณ์ระยะสั้น (ทรงศิริ, 2549) ซึ่งการพยากรณ์อนุกรมเวลาจะวัดสหสัมพันธ์ระหว่างค่าสังเกตที่ช่วงเวลาห่าง k ช่วงเวลา หรือเป็นการวัดความสัมพันธ์ระหว่าง Y_t และ Y_{t-k} อธิบายได้ว่า Y_t และ Y_{t-k} มีความสัมพันธ์ในทิศทางเดียวกันหรือในทิศทางตรงข้ามกัน

ขั้นตอนวิธีของอารีมา มีดังนี้

2.5.1) แบบจำลอง Auto Regressive เป็นรูปแบบที่แสดงว่าค่าสังเกต Y_t ถูกกำหนดจากค่าของ $Y_{t-1},\ldots, Y_{t-p}$ หรือค่าสังเกตุที่เกิดขึ้นก่อนหนา p โดยกระบวนการหรือระบบ AR(p) คือ กระบวนการหรือระบบ Auto Regressive ที่มีอันดับที่ p ซึ่งเขียนอยู่ในรูปสมการได้ดังนี้

AR (p) คือ $$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t$$

μ คือ ค่าคงที่ (Constant Term)

ϕ_j คือ พารามิเตอร์ตัวที่ j

ε_t คือ ความคลาดเคลื่อน ณ เวลา t

2.5.2) แบบจำลอง Moving Average (MA(q)) แบบจำลอง Moving Average (MA) เป็นรูปแบบที่แสดงว่าค่าสังเกต Y_t ถูกกำหนดจากค่าความคลาดเคลื่อน ε_{t-1},..., ε_{t-q} หรือค่าความคลาดเคลื่อนที่อยู่ก่อนหนา โดยกระบวนการ หรือระบบ MA(q) คือกระบวนการหรือระบบ Moving Average ที่มีอันดับ q ซึ่งเขียนในรูปของ MA(q) ได้ดังนี้

MA (q) คือ $$x_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}$$

μ คือ ค่าคงที่ (Constant Term)

θ_j คือ พารามิเตอร์เฉลี่ยเคลื่อนที่ตัวที่ j

ε_t คือ ความคลาดเคลื่อน ณ เวลา t

2.5.3) แบบจำลอง Auto Regressive Moving Average (ARMA(p,q)) แบบจำลอง Auto Regressive Moving Average (ARMA) เป็นแบบจำลองที่นำเอากระบวนการ Auto Regressive และ Moving Average มาใช้รวมกัน โดยกระบวนการหรือระบบ ARMA (p,q) คือกระบวนการหรือระบบ Auto Regressive ที่มีอันดับที่ p และ Moving Average ที่มีอันดับ q ซึ่งเขียนอยู่ในรูปสมการได้ดังนี้

แบบจำลอง ARMA (p,q)

$$y_t = \delta + \phi y_{t-1} + \phi y_{t-2} + \ldots + \phi y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - .. - \theta_q\varepsilon_{t-q}$$

Y_t คือ ค่าสังเกตในอนุกรมเวลา ณ เวลา t

p คือ อันดับของ Auto Regressive

q คือ อันดับของ Moving Average

δ คือ ค่าคงที่ (Constant Term)

t คือ เวลา

ϕ คือ พารามิเตอร์ของ Auto Regressive

θ คือ พารามิเตอร์ของ Moving Average t

ε_t คือ กระบวนการ white noise ซึ่งก็คือ คาความคลาดเคลื่อน ณ เวลา t

2.5.4) กระบวนการ Integrated (I(d)) กระบวนการ Integrated (I(d)) เป็นการหาผลต่างของอนุกรมเวลาระหว่างข้อมูล ณ ปัจจุบันกับข้อมูลถอยหลังไป d คาบเวลาโดย สาเหตุที่ต้องทำการหาผลต่างของอนุกรมเวลาเนื่องจากแบบจำลอง ARIMA ต้องใช้ในการ

วิเคราะห์ข้อมูลอนุกรมเวลาที่มีคุณสมบัติคงที่ (Stationary) เท่านั้นโดยในกรณีข้อมูลอนุกรรมเวลาที่ใช้ในการวิเคราะห์มีคุณสมบัติไม่คงที่ (Nonstationary) จะต้องทำการแปลงข้อมูลดังกล่าวให้เป็นข้อมูลที่มีคุณสมบัติคงที่ก่อน โดยการหาผลต่างของข้อมูลอนุกรมเวลาก่อนที่นำไปสร้างแบบจำลอง ARIMA ซึ่งโดยทั่วไปแล้วว่าต้องหาผลต่างอันดับที่ d สามารถเขียนในรูปของ I (d) ได้ดังนี้

I(d) คือ $\Delta_d x_t = \Delta_{d-1}(x_t - x_{t-1})$ หรือ $(1-B)^d x_t$

ε_t คือ พจน์ความคลาดเคลื่อน ณ เวลา t

$(1-B)^d \chi_t$ คือ ผลต่างอันดับที่ d

B คือ Backward shift operation

2.5.5) แบบจำลอง Autoregressive integrated moving average model (ARIMA) จากรายละเอียดต่าง ๆ ที่กล่าวในข้างต้าน ถ้านำแบบจำลอง Auto Regressive แบบจำลอง Moving Average และกระบวนการ Integrated มาพิจารณารวมกันสามารถนำมากำหนดเป็นรูปแบบทั่วไปของแบบจำลอง ARIMA ที่ใช้ในการประมาณการคือแบบจำลอง ARIMA (p,d,q)

$$\Delta_d y_t = \delta + \phi\Delta_d y_{t-1} + \phi\Delta_d y_{t-2} + \ldots + \phi\Delta_d y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q}$$

Y_t คือ ค่าสังเกตในอนุกรมเวลา ณ เวลา t

d คือ จำนวนครั้งของการหาผลต่างเพื่อให้อนุกรมเวลามีคุณสมบัติคงที่ (Stationary)

p คือ อันดับของ Autoregressive

q คือ อันดับของ Moving Average

δ คือ ค่าคงที่ (Constant Term)

t คือ เวลา d

Δ^d คือ ผลต่างอันดับที่ d

ϕ_p คือ พารามิเตอร์ของ Auto Regressive

θ_p คือ พารามิเตอร์ของ Moving Average t

ε_t คือ กระบวนการ white noise ซึ่งก็คือ คาความคลาดเคลื่อน ณ เวลา t ภายใต้ข้อสมมติท่าว่าความคลาดเคลื่อนที่คนละเวลาเป็นตัวแปรสุมที่เป็นอิสระต่อกันโดยมีการแจกแจงปกติที่มีค่าเฉลี่ยเป็นศูนย์ละความแปรปรวนคงที่

นอกจากส่วนประกอบต่าง ๆ ของแบบจำลอง ARIMA ที่กล่าวในข้างต้นนั้น ก่อนที่จะสร้างแบบจำลองจำเป็นต้องเข้าใจในเรื่องของ Autocorrelation Function และ Partial Autocorrelation Function

2.5.6) Autocorrelation Function (ACF) Autocorrelation Function เป็นฟังก์ชั่นของการวัดสหสัมพันธ์ระหว่างข้อมูล ณ เวลา t ($\mathbf{X_t}$) และ ข้อมูล ณ เวลา t-k ($\mathbf{X_{t-k}}$) ของช่วงเวลาห่างกัน k หน่วย ซึ่งแทนด้วยสัญลักษณ์ ρ_k หรือ ρ_k ในกรณีสหสัมพันธ์ในตัวเองของตัวอย่าง ซึ่งสามารถคำนวณได้ดังนี้

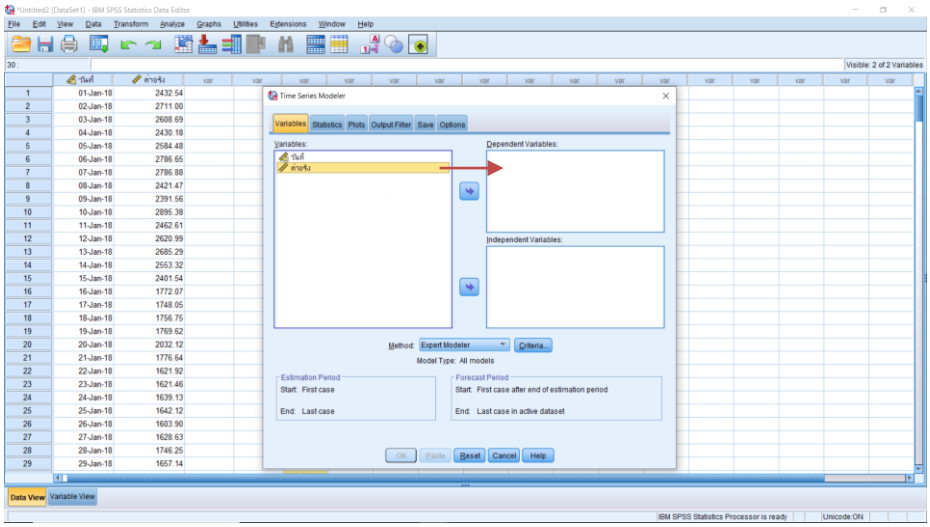
$$\rho_k \text{ หรือ } r_k = \frac{\sum_{t=k+1}^{n} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2}$$

2.5.7) Partial Autocorrelation Function (PACF) การพิจารณาสหสัมพันธ์ระหว่างตัวแปร $\mathbf{X_t}$ กับ $\mathbf{X_{t-k}}$ อาจเป็นไปได้ว่าสหสัมพันธ์ดังกล่าว เป็นผลเนื่องมาจากสหสัมพันธ์ระหว่าง 2 ตัวแปรนี้กับตัวแปร $\mathbf{X_{t-1}}$,..., $\mathbf{X_{t-k+1}}$ ดังนั้นเพื่อที่จะได้สหสัมพันธ์ระหว่าง $\mathbf{X_t}$ กับ $\mathbf{X_{t-k}}$ ที่ได้ขจัดความเกี่ยวข้องระหว่างตัวแปรทั้งสองตัวนี้กับตัวแปร $\mathbf{X_{t-1}}$,..., $\mathbf{X_{t-k+1}}$ ดังกล่าว จึงต องทำการวัดสหสัมพันธ์ของทั้งสองตัวแปรในรูปแบบของการสหสัมพันธ์แบบมีเงื่อนไข Corr ($\mathbf{X_1}$, $\mathbf{X_{t-k}}$|$\mathbf{X_{t-1}}$,..., $\mathbf{X_{t-k+1}}$) ซึ่งเรียกว่า Partial Autocorrelation โดยแทนด้วยสัญลักษณ์ ϕ_{kk} แต่ถ้านำสัมพันธ์ในตัวเองบางส่วนมาพิจารณาในรูปแบบฟังก์ชั่นจะเรียกว่าPartial Autocorrelation Function (PACF) ซึ่ง ϕ_{kk} สามารถคำนวณได้ดังนี้

$$\phi_{kk} = \frac{Cov[(x_t - \hat{x}_t), (x_{t-k} - \hat{x}_{t-k})]}{\sqrt{Var(x_t - \hat{x}_t)}\sqrt{Var(x_{t-k} - \hat{x}_{t-k})}}$$

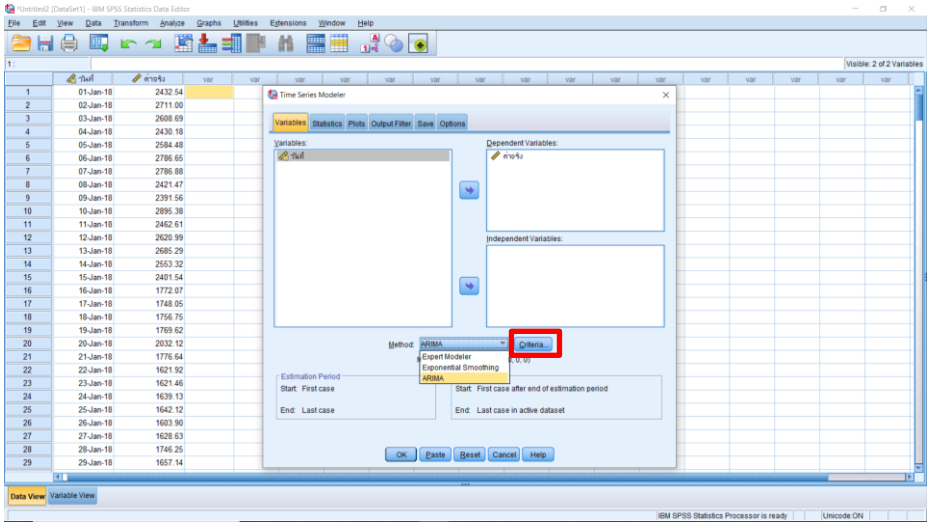
3) การตั้งค่าการพยากรณ์ด้วย ARIMA

3.1) เลือกค่าจริงใส่ในช่อง Dependent Variables



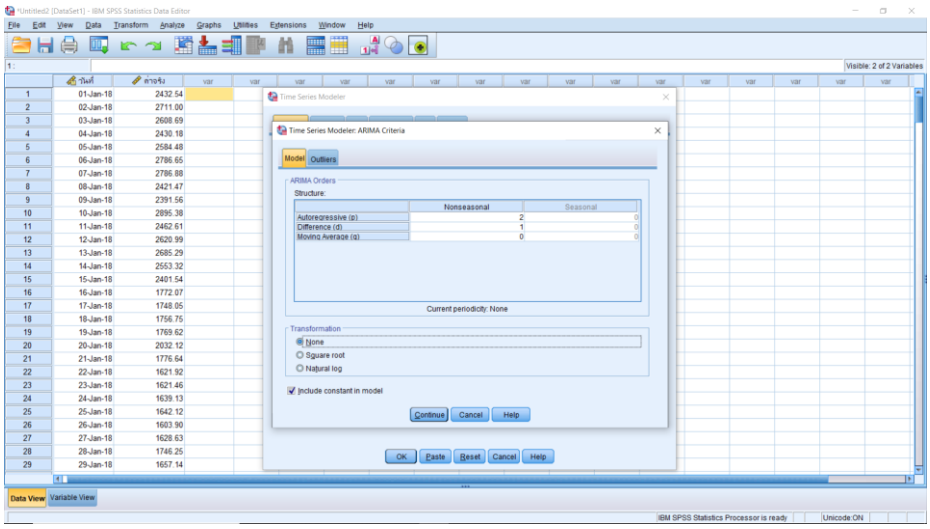
ภาพที่ 3.22 การตั้งค่าการพยากรณ์ด้วยวิธี ARIMA (1)

3.2) เลือก Method เป็น ARIMA และเลือก Criteria



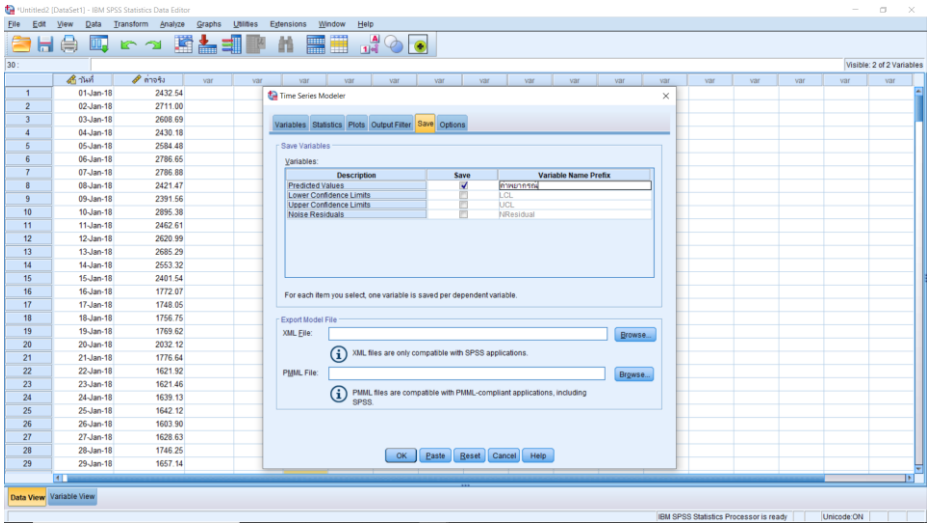
ภาพที่ 3.23 การตั้งค่าการพยากรณ์ด้วยวิธี ARIMA (2)

3.3) กำหนดค่าที่เหมาะสม



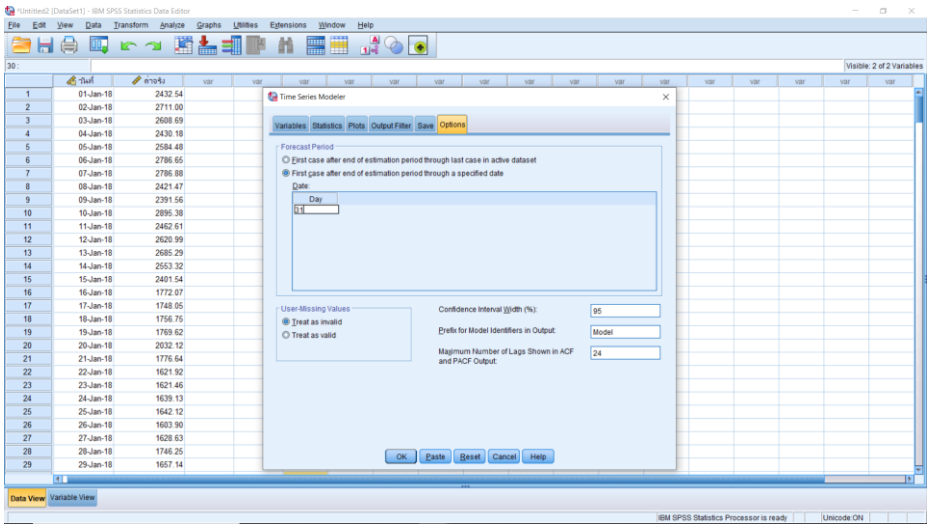
ภาพที่ 3.24 การตั้งค่าการพยากรณ์ด้วยวิธี ARIMA (3)

3.4) ทำการเลือกบันทึก Predicted Values และตั้งชื่อ



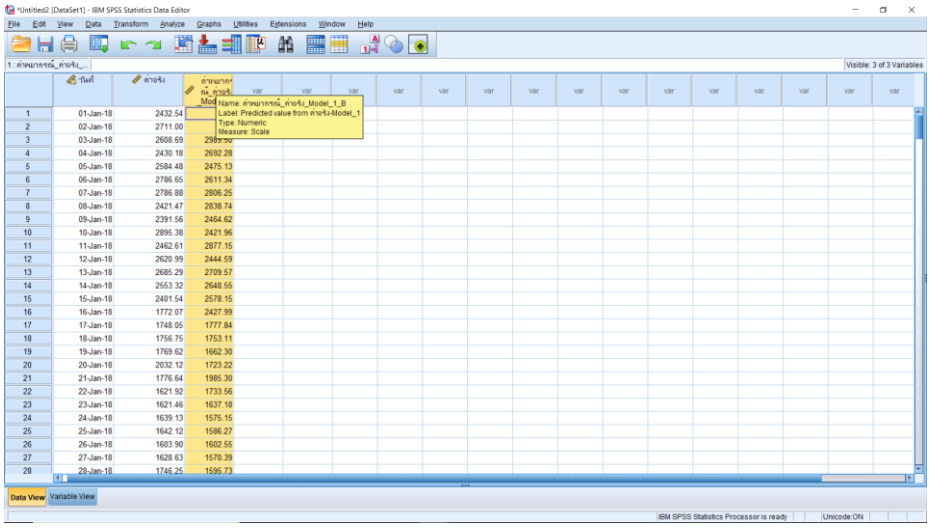
ภาพที่ 3.25 การตั้งค่าการพยากรณ์ด้วยวิธี ARIMA (4)

3.5) ใส่จำนวนวันเท่ากับ 31 วันเนื่องจากเดือนธันวาคมมี 31 วัน

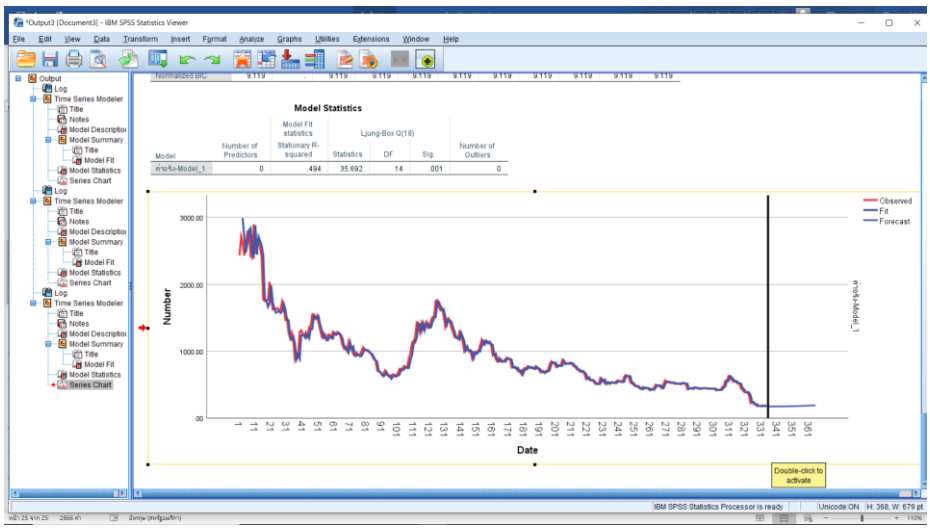


ภาพที่ 3.26 การตั้งค่าการพยากรณ์ด้วยวิธี ARIMA (5)

3.6) ค่าพยากรณ์ด้วยวิธี ARIMA ล่วงหน้า 31 วัน



ภาพที่ 3.27 แสดงค่าพยากรณ์ด้วยวิธี ARIMAล่วงหน้า 31 วัน



ภาพที่ 3.28 แสดงผลการพยากรณ์ด้วยวิธี ARIMA ล่วงหน้า 31 วัน

3.1.5 Evaluation ผู้วิเคราะห์ข้อมูลได้ผลการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้าไมน์ นิ่งจากการสร้าง โมเดล Time Series สามารถวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรกว่า ข้อมูลในวิธีมีความน่าเชื่อถือมากกว่า โดยการนำข้อมูลการพยากรณ์ล่วงหน้าของทั้ง 2 วิธีมาหาค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย และนำมาเปรียบเทียบผลที่ได้

เกณฑ์การเปรียบเทียบประสิทธิภาพ วิธีการพยากรณ์

ในการศึกษาครั้งนี้ ผู้วิเคราะห์ ได้เปรียบเทียบวิธีการพยากรณ์ 2 วิธีดังกล่าว โดยพิจารณาจากค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error : MAPE) ซึ่งการวัดความคลาดเคลื่อนระหว่างค่า จริงและค่าพยากรณ์ ค่า MAPE นี้ไม่มีหน่วย และมีค่าเป็นบวกเสมอ

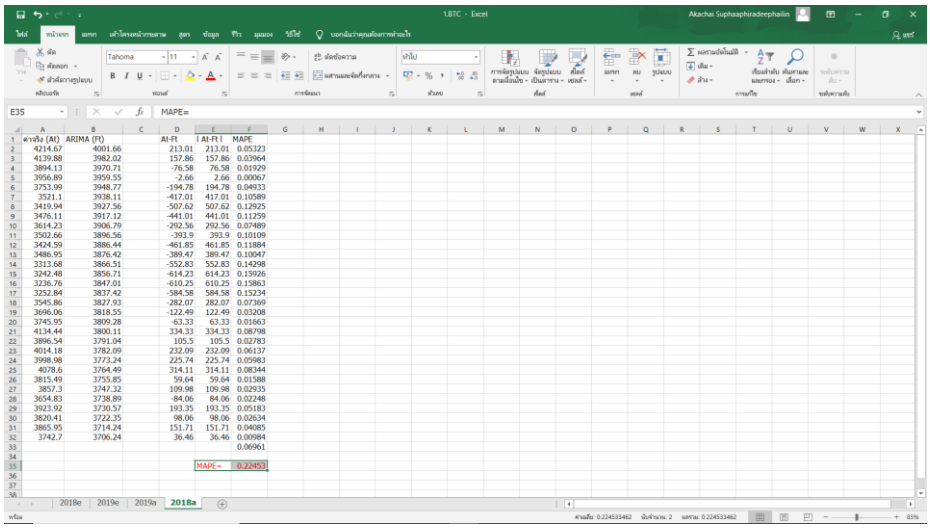
$$MAPE = \frac{\sum \frac{|A-F|}{A} \times 100}{N}$$

โดย F หมายถึงค่าพยำกรณ์ในงวดที่ 1

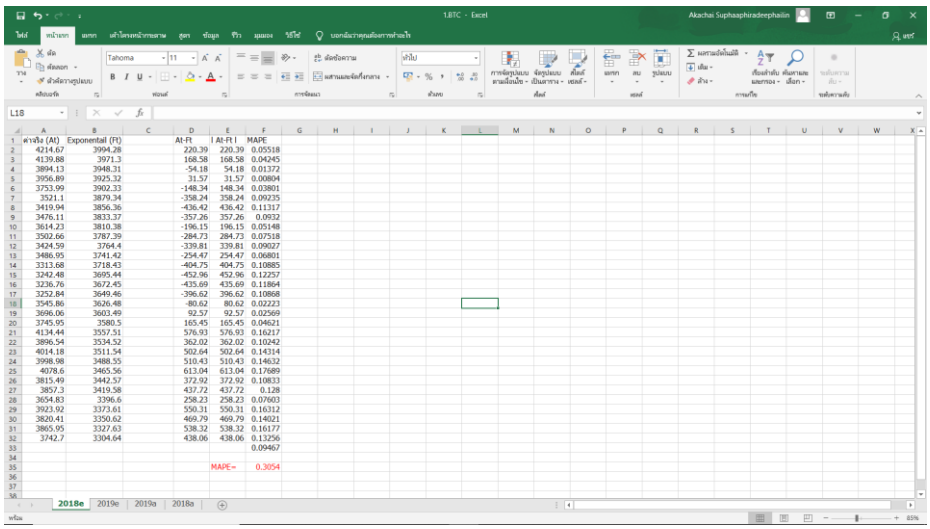
A หมายถึงยอดที่เกิดขึ้นจริงในช่วงเวลา t

N หมายถึง จำนวนข้อมูล

ผู้วิเคราะห์ใช้โปรแกรม Excel สร้างตารางและคอลัมน์ นำข้อมูลของเดือนที่ 12 ทั้งข้อมูลจริงและค่าพยากรณ์ล่วงหน้า 1 เดือนจากวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (EMA) และวิธีอารีมา (Arima) มาคำนวณหาค่า MAPE เพื่อเปรียบเทียบวิธีพยากรณ์ที่แม่นยำสำหรับการการพายากรณ์ล่วงหน้า



ภาพที่ 3.29 แสดงค่า MAPE จากการพยากรณ์ด้วยวิธี ARIMA

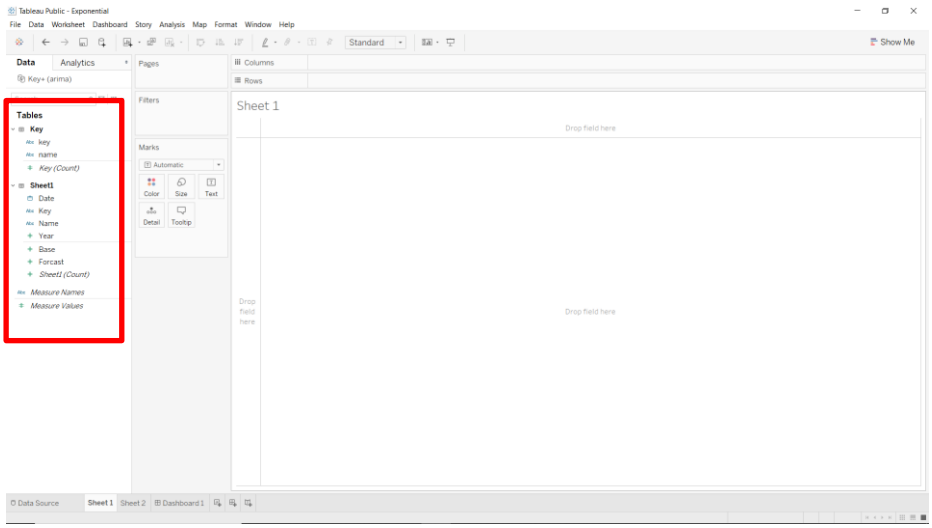


ภาพที่ 3.30 แสดงค่า MAPE จากการพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล

จากข้อมูลข้างต้นผู้วิเคราะห์ได้ทำการหาค่า MAPE ของวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล และวิธีอารีมาในปี 2562 ของหุ้น Bitcoin(BTC)

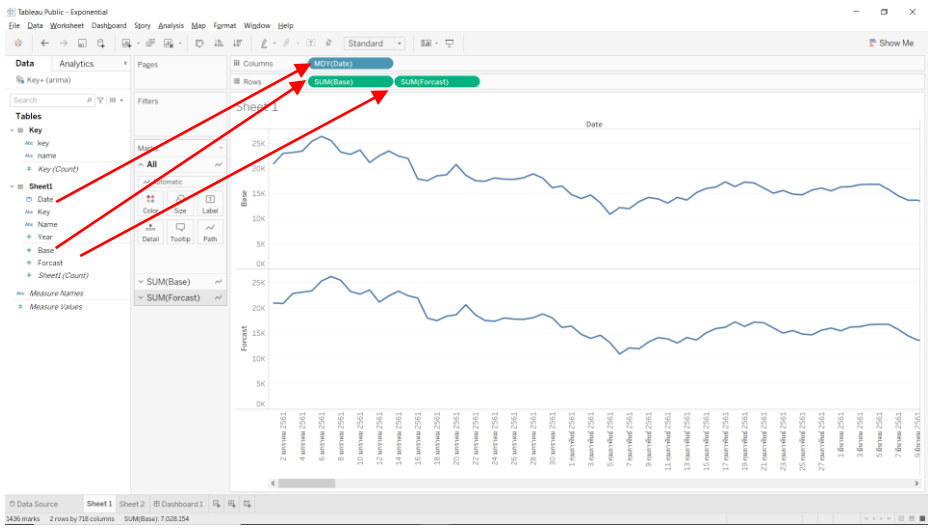
3.1.6 การนำผลลัพธ์ไปใช้งาน (Deployment) ผู้วิเคราะห์ข้อมูลนำผลองค์ความรู้ที่ได้เหล่านี้ไปนำเสนอแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public โดยดำเนินการตามขั้นตอนดังต่อไป

1.)ข้อมูลมาจัดแสดงให้เป็นกราฟ การทำงาน ข้อมูลที่ Import เข้ามาจะอยู่แถบด้านซ้าย



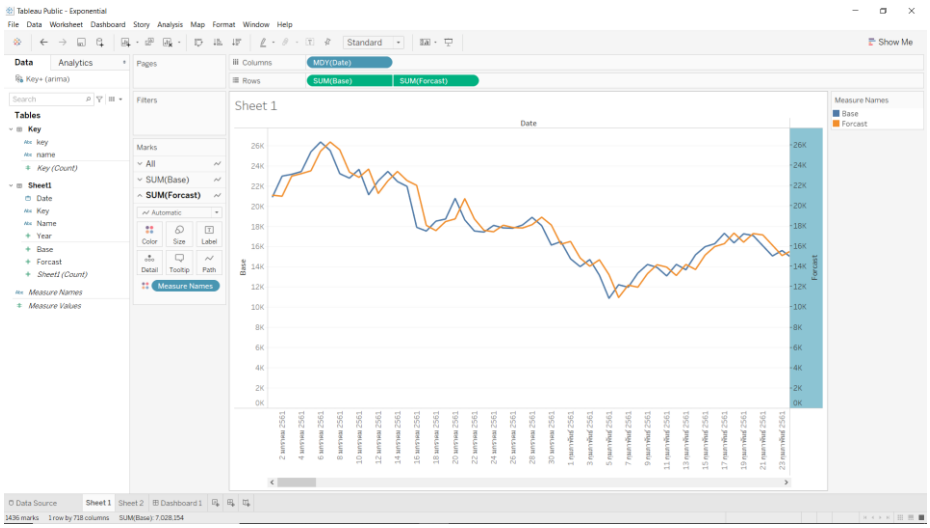
ภาพที่ 3.31 แสดงพื้นที่การทำงานของโปรแกรม

2.) ทำการดึงข้อมูลค่าจริง และค่าพยากรณ์มาจัดแสดงในพื้นที่จัดแสดงกราฟ



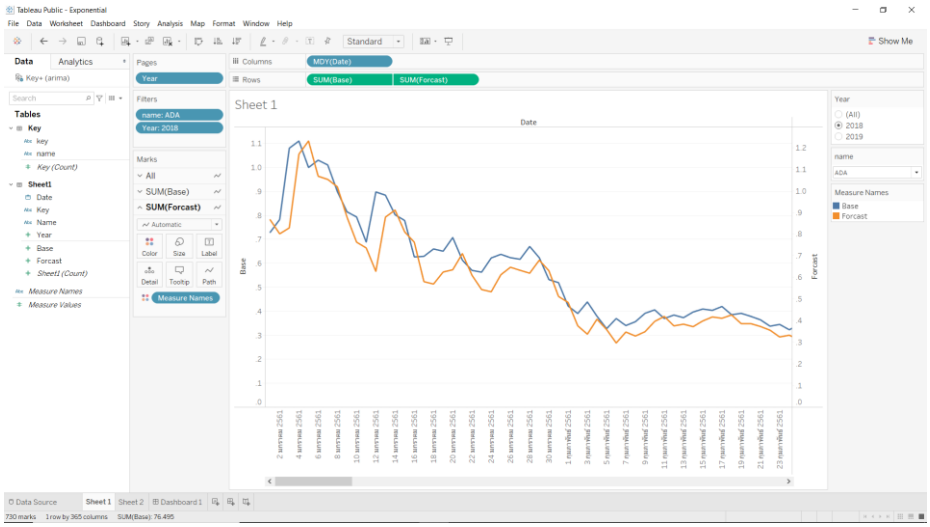
ภาพที่ 3.32 แสดงข้อมูลจริงและข้อมูลพยากรณ์

3.) ทำการเปลี่ยนรูปแบบการจัดแสดงให้อยู่ในกราฟเดียวกัน เพื่อง่ายต่อการเปรียบเทียบ



ภาพที่ 3.33 แสดงกราฟในรูปแบบเปรียบเทียบ

4.) ทำการเพิ่มตัวกรองรายการเพื่อสามารถเรียกดูเฉพาะรายการที่สนใจได้



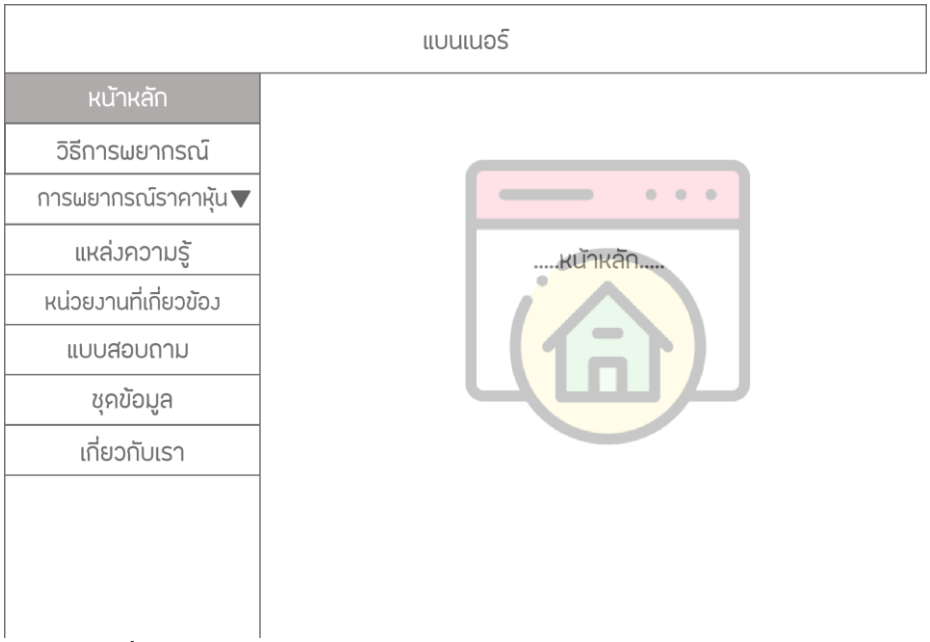
ภาพที่ 3.34 การเพิ่มตัวกรองรายการ

3.2 การออกแบบและพัฒนาเว็บไซต์

ผู้วิเคราะห์ได้วางแผนจัดลำดับเนื้อหาของเว็บไซต์ออกเป็นหมวดหมู่ เพื่อจัดรูปแบบโครงสร้างในการจัดวางเมนูบนหน้าเว็บเพจ ทำให้เห็นรูปแบบโครงสร้างทั้งหมดของเว็บไซต์ และการจัดระเบียบข้อมูลที่ชัดเจน ให้แยกย่อยเนื้อหาออกเป็นส่วนต่าง ๆ ที่อยู่ในหมวดหมู่เดียวกัน ทำให้การใช้งานเว็บไซต์นั้น และเข้าถึงเนื้อหาภายในเว็บไซต์นั้นง่าย

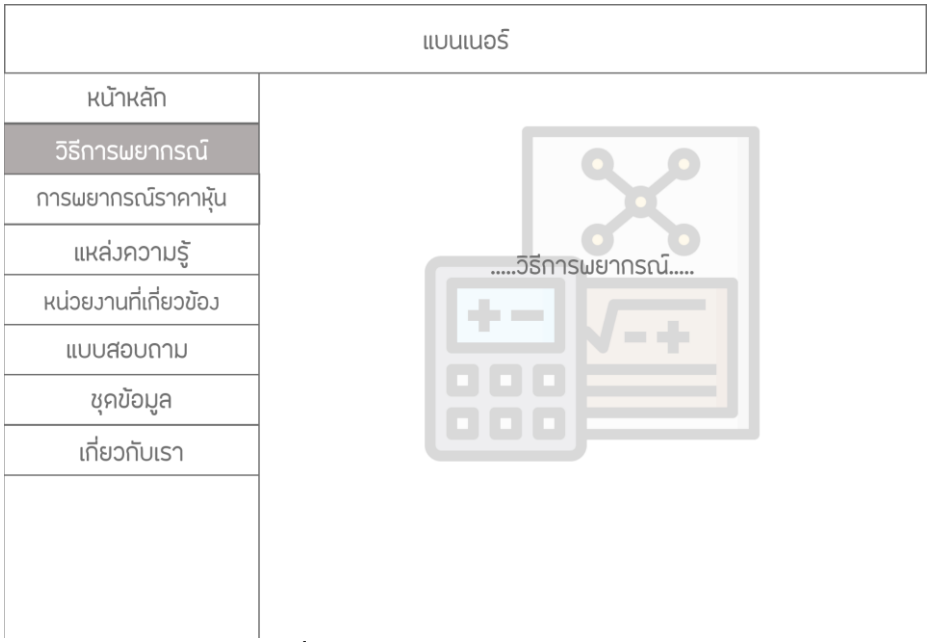
3.2.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

1) หน้าหลักของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์



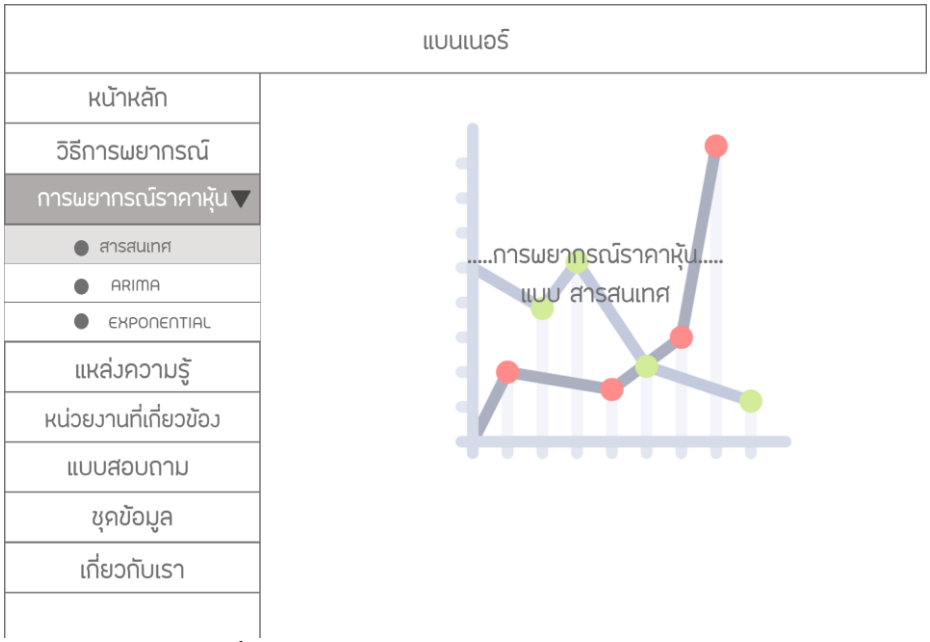
ภาพที่ 3.35 แสดงหน้าหลักของเว็บไซต์ แสดงเมนูต่าง ๆ ของเว็บไซต์

2) หน้าแสดงวิธีในการพยากรณ์



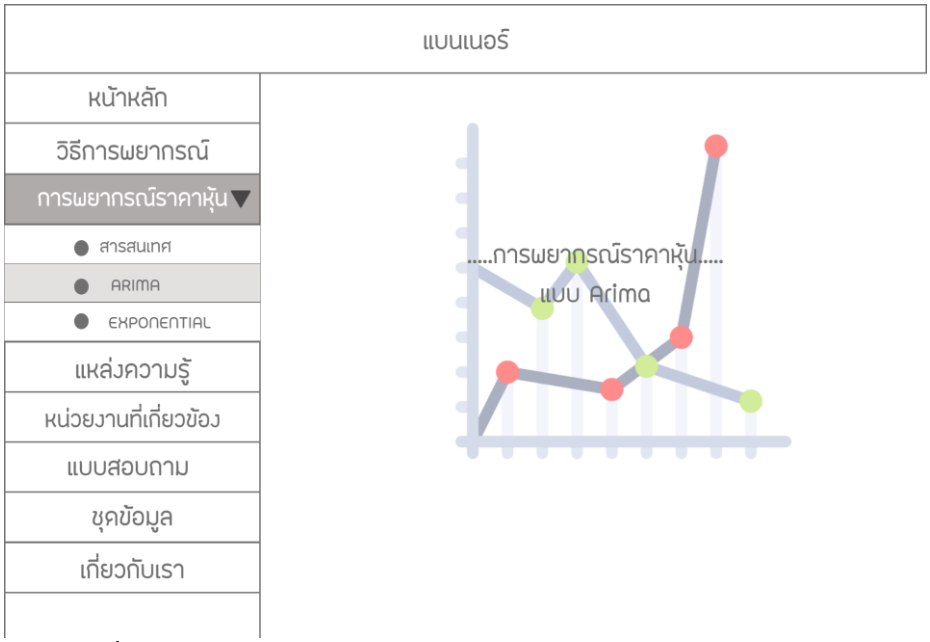
ภาพที่ 3.36 หน้าแสดงวิธีการพยากรณ์

3) หน้าแสดงข้อมูลการพยากรณ์ราคาหุ้นแบบสารสนเทศหุ้นคริปโตเคอร์เรนซี



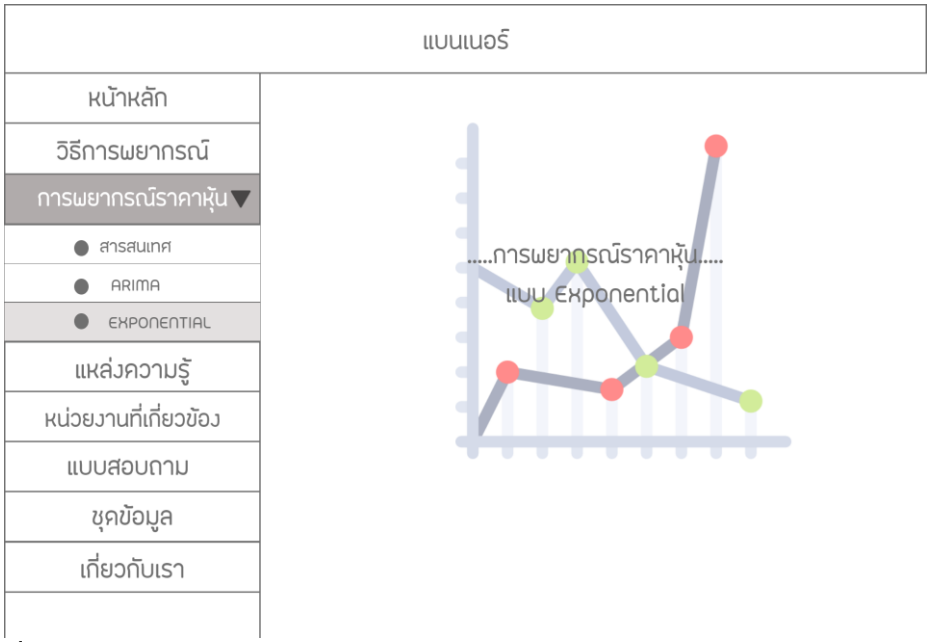
ภาพที่ 3.37 หน้าแสดงข้อมูลการพยากรณ์ราคาหุ้นสารสนเทศ

4) หน้าแสดงข้อมูลการพยากรณ์ราคาหุ้นแบบอารีมา (Arima)



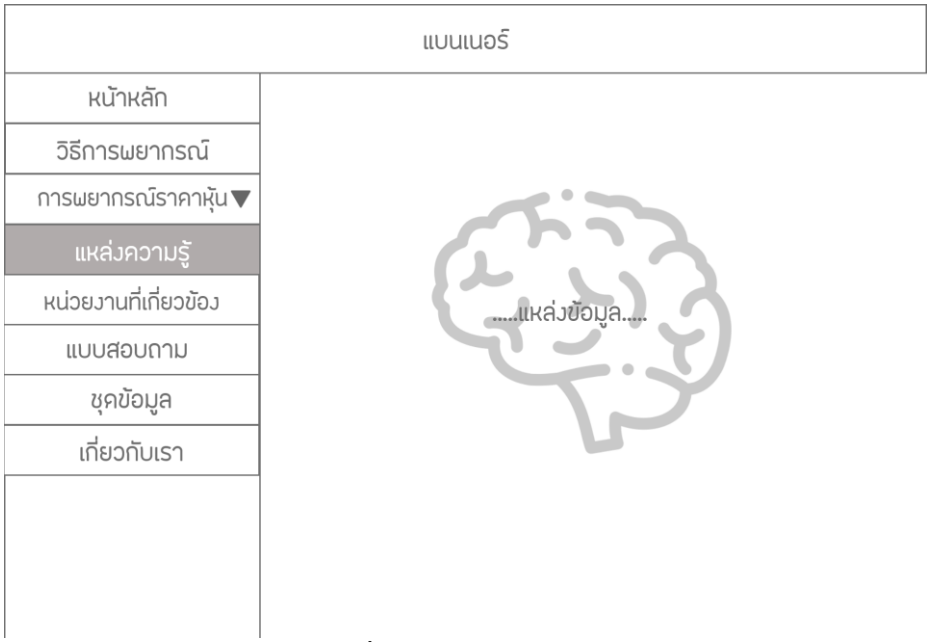
ภาพที่ 3.38 หน้าแสดงข้อมูลการพยากรณ์ราคาหุ้นแบบอารีมา (Arima)

5) หน้าแสดงข้อมูลการพยากรณ์ราคาหุ้นแบบเอ็กซ์โปเนนเชียล (Exponential)



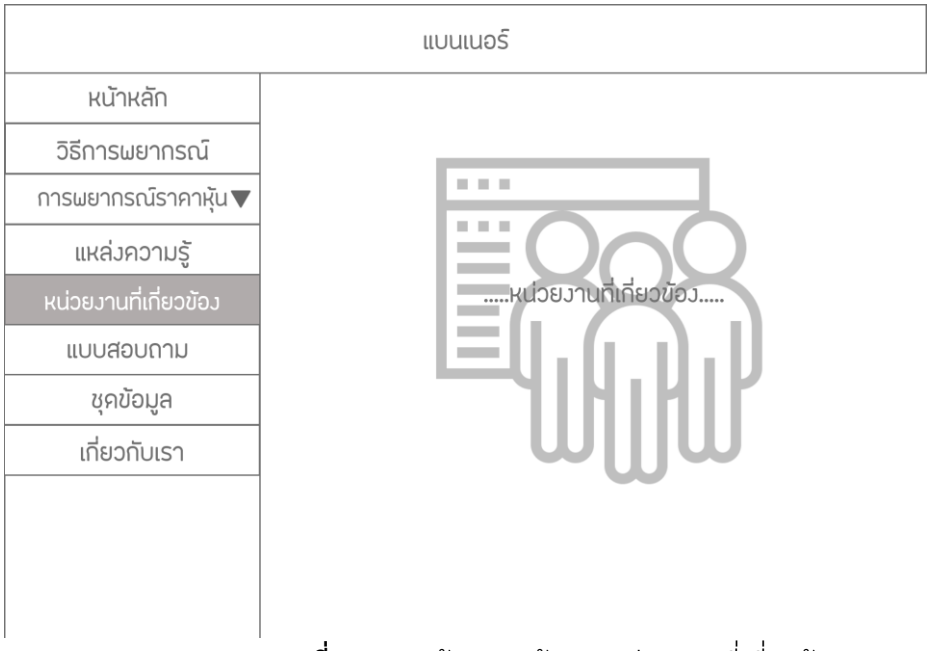
ภาพที่ 3.39 หน้าแสดงข้อมูลการพยากรณ์ราคาหุ้นแบบเอกซ์โพเนนเชียล (Exponential)

6) หน้าแสดงเว็บไซต์แหล่งความรู้



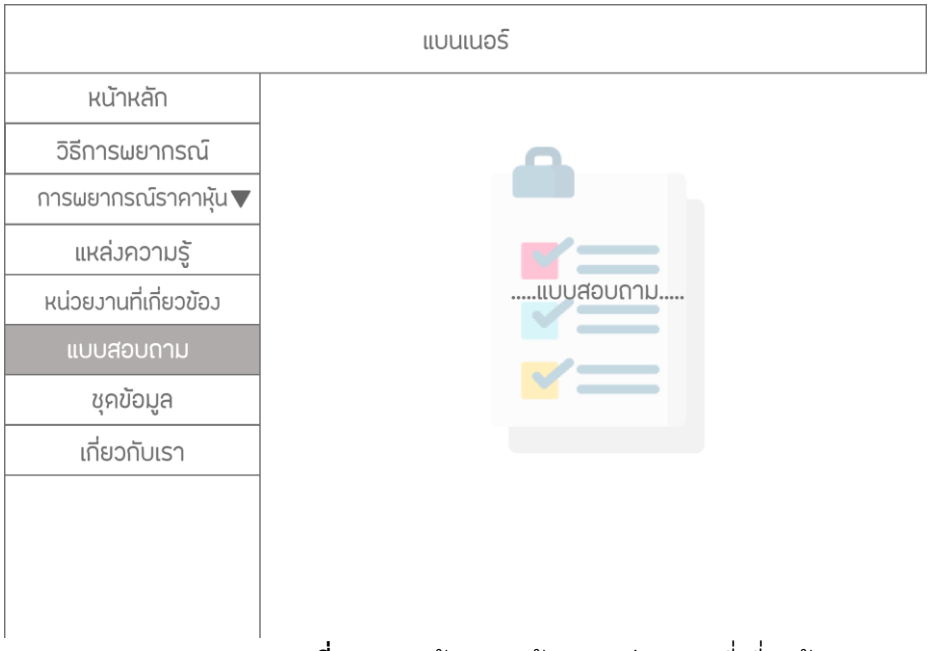
ภาพที่ 3.40 หน้าแสดงข้อมูลแหล่งความรู้

7) หน้าแสดงเว็บไซต์หน่วยงานที่เกี่ยวข้อง



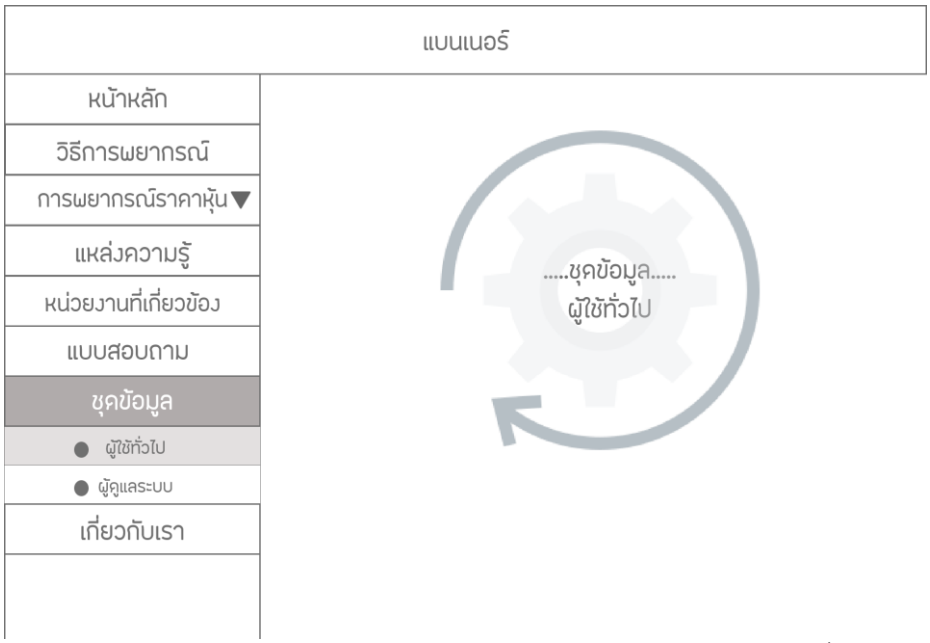
ภาพที่ 3.41 หน้าแสดงข้อมูลหน่วยงานที่เกี่ยวข้อง

8) หน้าแสดงเว็บไซต์แบบสอบถาม



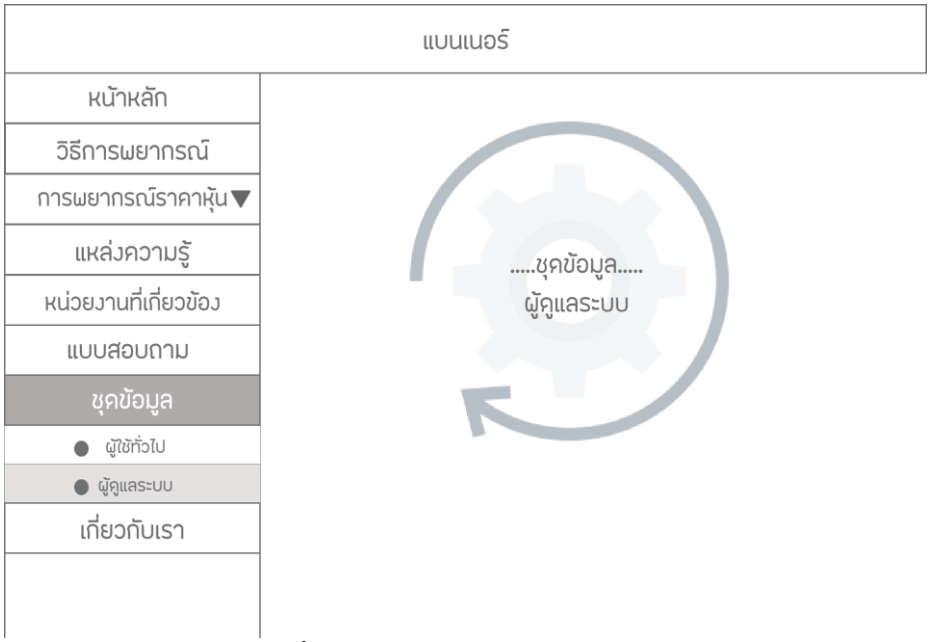
ภาพที่ 3.42 หน้าแสดงข้อมูลหน่วยงานที่เกี่ยวข้อง

9) หน้าแสดงเว็บไซต์ชุดข้อมูลของผู้ใช้งาน



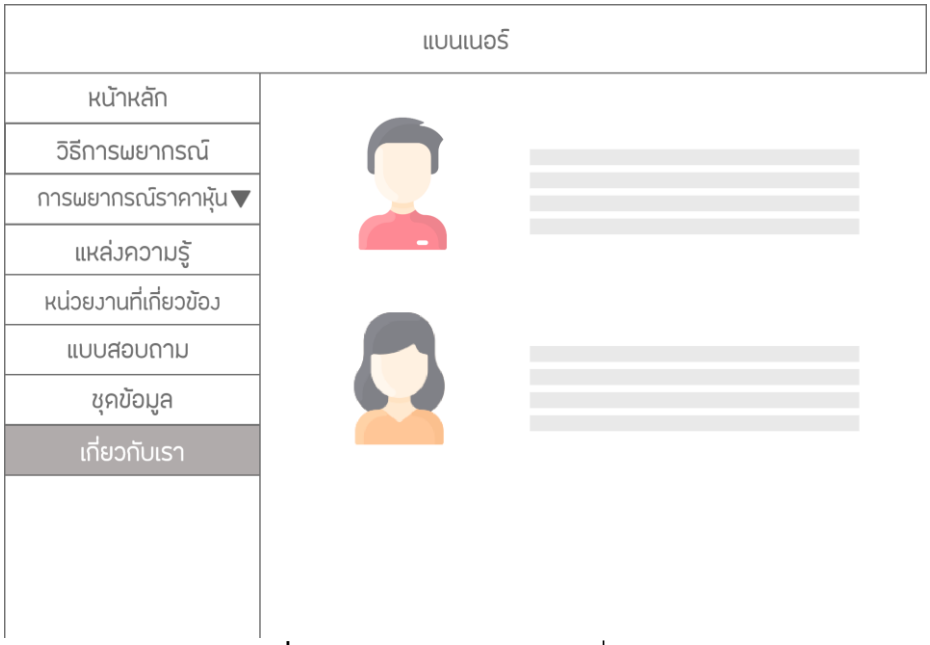
ภาพที่ 3.43 หน้าแสดงข้อมูลชุดข้อมูลของผู้ใช้งานทั่วไป

10) หน้าแสดงเว็บไซต์ชุดข้อมูลของผู้ดูแลระบบ



ภาพที่ 3.44 หน้าแสดงข้อมูลชุดข้อมูลของผู้ดูแลระบบ

11) หน้าแสดงเว็บไซต์เกี่ยวกับเรา



ภาพที่ 3.45 หน้าแสดงเว็บไซต์เกี่ยวกับเรา

3.3 บทสรุป

จากวิธีการดำเนินงานโครงงานในข้างต้น ผู้วิเคราะห์ได้แสดงวิธีในการจัดเก็บข้อมูลข้อมูลราคาหุ้นคริปโตเคอร์เรนซี ด้วยขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP–DM มาใช้ในการรวบรวมข้อมูล และวิเคราะห์ข้อมูลที่เกี่ยวข้องกับงาน ซึ่งเป็นหลักในการจัดทำเหมืองข้อมูลเพื่อวิเคราะห์และใช้ประโยชน์ในทางธุรกิจ เพื่อได้สารสนเทศของงานอย่างเพียงพอในการนำไปใช้ประโยชน์ รวมถึงการสร้างโมเดล Time Series ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล และอารีมาจากโปรแกรมที่ใช้เลือกทำเหมืองข้อมูลเพื่อใช้นำเสนอ คือ โปรแกรม IBM SPSS ในการสร้างโมเดล Time Series และประเมินประสิทธิภาพของโมเดล ด้วยค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย ซึ่งจะเลือกค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยที่ต่ำที่สุด ผู้วิเคราะห์ได้จัดทำข้อมูลสารสนเทศมาทการแสดงผลแบบ visualization โดยใช้โปรแกรม Tableau Public และออกแบบ Wireframe ของเว็บไซต์ที่จะเผยแพร่บนเว็บไซต์