



มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

วิชา CSS 341 Introduction to Data Science

สำหรับ นศ. ภาควิชาคณิตศาสตร์

ข้อสอบครั้งที่ 2 ปีการศึกษา 2566

วันอังคารที่ 24 กันยายน 2566 เริ่มต้นเวลา 12:00 น.

คำแนะนำและคำสั่ง

ข้อสอบมี 2 ส่วน (Parts) รวม 9 ข้อ คิดเป็น $29 + 31 = 60$ คะแนน ให้ทำทุกข้อ โดยที่....

- เขียน Python 3.8 หรือ 3.9 ในรูปแบบ ipynb **ไฟล์เดียวต่อเนื่อง** ไปเลยทุกข้อ
- เขียน Markdown ให้เหมาะสมชัดเจน นั่นคือ สำหรับแต่ละข้อให้ ควรมี โจทย์ (ย่อมาสั้นๆ) ตามด้วยโค้ด และตามด้วย(การวิเคราะห์) ผลลัพธ์เพื่อสรุปตอบ เรียงลำดับข้อไป
- ส่วนต้นสุด**ของ ipynb file ที่ส่ง**จะต้องมี** student ID และตามด้วยรายชื่อของสมาชิกทุกคนชัดเจน (เขียนเป็น Markdown ไว้) **หากไม่มีจะถูกหัก 3 คะแนน**
- เพื่อความสะดวกในการตรวจหลายกลุ่ม ในทำข้อสอบให้กำหนด data files ทุกไฟล์อยู่ใน folder `..` คือเหนือขึ้นไป 1 ชั้นกับ folder ที่บรรจุ ipynb ของท่าน ดังนั้น เวลาอ่าน data file ให้กำหนด path เป็น `'../<filename>'` เมื่อ `<filename>` คือชื่อไฟล์ที่อ่านเข้ามา หากผิดกติกานี้จะถูกหัก 2 แต้มต่อการอ่าน data file 1 ครั้ง
- การส่ง ให้ส่งขึ้น LEB2 ด้วย file ipynb เท่านั้น. ไม่ต้องส่ง data file มา เพราะมีอยู่แล้ว กลุ่มที่ส่งสายเกินกำหนดจะถูกหัก 5 นาทีละ 1 คะแนน
- ตอนท้ายข้อสอบนี้จะมีตัวอย่างบางส่วนของไฟล์ข้อมูลต่างๆ ที่ต้องใช้ทั้ง 7 ไฟล์

Name	Date modified
covid19	12/9/2566 16:18
gdp_per_capita_growth	12/9/2566 14:38
accidents	30/8/2566 10:59
sales	30/8/2566 9:33
shops	30/8/2566 9:22
item_categories	30/8/2566 9:11
items	13/9/2563 10:22
team 3	12/10/2566 11:10
team 2	12/10/2566 11:10
your group folder	12/10/2566 11:08

Part 1 (29 คะแนน)

ไฟล์ที่ใช้ได้แก่ sales.csv, shops.csv, items.csv และ item_categories.csv

- ในที่นี้ เราจะไม่สนใจการคืนสินค้า ดังนั้น ให้ตัดรายการ ข้อมูล sales.csv ที่มีค่า item_cnt_day หรือ item_price น้อยกว่า 0 และแสดงสรุปข้อมูลให้เห็นในภาพรวมว่าตัดออกไปแล้ว (1 คะแนน)
- ทุกร้านค้ารวมกันขายสินค้าไปได้กี่ชนิดสินค้า (1 คะแนน)
- สินค้าชนิดใดที่ขายได้มากที่สุดในแต่ละร้าน โดยแสดงชื่อร้าน ชื่อสินค้านั้น และจำนวนที่ร้านนั้นขายสินค้านั้นได้ โดยแสดงมาเพียง 10 ร้านค้าที่มีจำนวนชิ้นที่ขายได้สูงสุดพอ (4 คะแนน) (ทุกข้อไม่ต้องมีเส้นตาราง)

Shop Name	Most sold item name	Number of items sold

4. แสดงค่าเฉลี่ย, ค่ามัธยฐาน, ค่าสูงสุดและค่าต่ำสุด ของราคาสินค้าของแต่ละร้านค้า โดยเรียงลำดับตามชื่อของร้านค้าจาก A ไป Z รวมทั้งจำนวนสินค้าที่ร้านนั้นมีขายด้วย แสดงผลลัพธ์ดังตารางต่อไปนี้ (4 คะแนน)

Shop Name	Number of items	Average	Median	Highest	Lowest

5. ยอดขายคำนวณได้จากราคาสินค้า * จำนวนที่ขายได้ ในที่นี้ ให้นำเสนอข้อมูลในรูปแบบตาราง เพื่อบรรยายงานสรุปดังตารางต่อไปนี้ โดยที่

5.1. แต่ละร้านค้าขายสินค้าแต่ละหมวด (category) รวมได้มูลค่าเท่าใด (คอลัมน์ที่ 3) (5 คะแนน)

5.2. โดยให้เรียงลำดับด้วยชื่อร้านค้า (Shop Name) (จากน้อยไปมาก) แล้วยอดขายของแต่ละหมวดสินค้า (คอลัมน์ที่ 3) จากมากไปน้อย (3 คะแนน)

Shop Name	Product category	Sales
Home Depot	Desks	6,300.00
	Stationery	2,497.12
Top Supermarket	Meats	34,502.34
	Vegetables	1,278.90

6. ในข้อนี้ ให้พิจารณาร้านค้าที่มียอดขายรวมสูงสุดเท่านั้น

6.1. ให้แสดงและวิเคราะห์ยอดขายของร้านโดยแยกเป็นรายเดือน และรายวันของสัปดาห์ (Day of week)

(5 คะแนน)

Month	Sales
January	
...	
July	
...	
December	

Day of week	Sales
Sunday	
...	
Saturday	

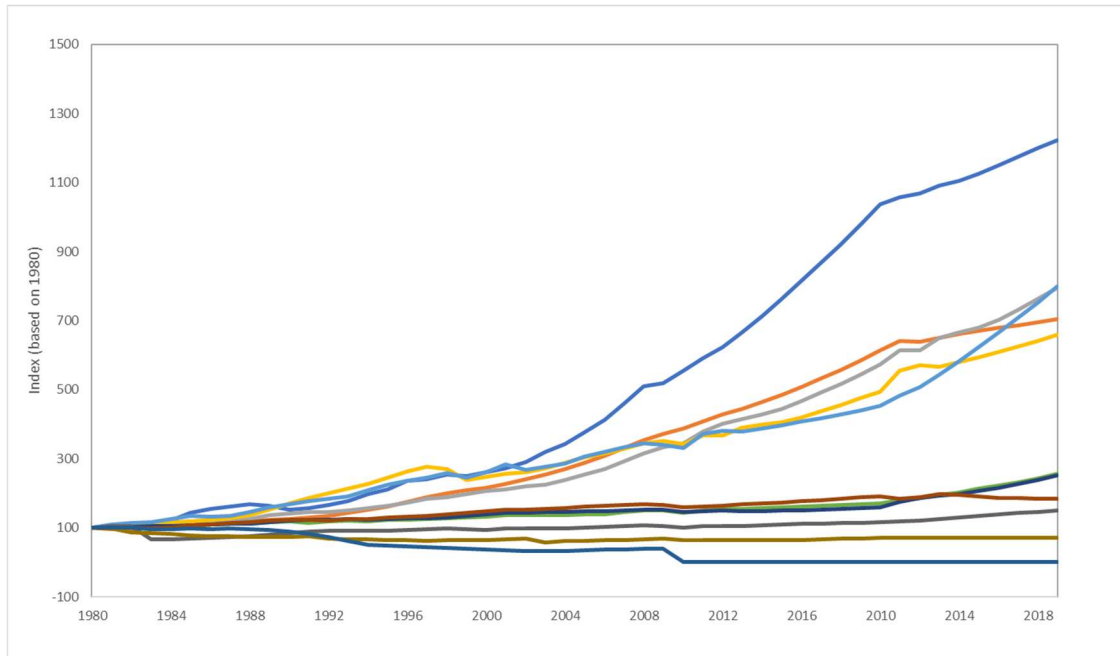
6.2. นำข้อมูลในข้อ 6.1 มาแสดงเป็นกราฟ (Chart) อย่างเหมาะสม (2 คะแนน)

6.3. จากนั้นให้แสดงกราฟ (Chart) เพื่อให้เห็นว่าแต่ละวันของสัปดาห์ (Day of week) ขายได้คิดเป็นสัดส่วนเท่าใด

(2 คะแนน)

Part 2 (6 + 11 + 14 = 31 คะแนน)

7. จากข้อมูล `gdp_per_capita_growth.csv` หรือ อัตราการเติบโตของผลิตภัณฑ์มวลรวมต่อประชากร ที่ให้มาของ 220 ประเทศตั้งแต่ปี 1980 ถึง 2019 ให้นำเสนอ มูลค่าผลิตภัณฑ์มวลรวมต่อประชากร ของ China, Vietnam, India, Thailand, Singapore, Germany, France, USA, Japan, Madagascar, และ Congo โดยกำหนดค่าตั้งต้นเท่ากับเป็น 100 ที่ปีฐาน 1980 และให้นำเสนอเป็นกราฟที่สื่อความหมายคล้ายกับรูปต่อไปนี้ (แต่ควรจะดีกว่า) (6 คะแนน)





8. พิจารณาข้อมูลยอดผู้เสียชีวิตสะสม `covid19.csv`
- 8.1. ในกลุ่มประเทศที่มีผู้เสียชีวิต (เฉพาะวันล่าสุด) มากที่สุด 20 ประเทศ ให้นำเสนอ สัดส่วน ของจำนวนผู้ป่วยในแต่ละประเทศเมื่อเทียบกับทั้งหมด (20 ประเทศนี้) ทั้งนี้ให้รวมประเทศลำดับที่ 11-20 เป็น อื่นๆ (Others) (4 คะแนน)
- 8.2. พิจารณาเฉพาะประเทศแคนาดา (รวมกัน 9 วันของเดือนล่าสุด Mar 2023) ให้สร้างกราฟแสดงจำนวนผู้เสียชีวิตสะสมเรียงลำดับจากมากไปน้อย โดยให้เลือกแสดงเฉพาะรัฐที่มีผู้เสียชีวิตหลักหมื่นขึ้นไปเท่านั้น (4 คะแนน)
- 8.3. หาจำนวนรวมผู้เสียชีวิตที่เกิดขึ้นในแต่ละเดือน ของแต่ละปี และสร้าง Bubble Chart โดยกำหนดให้แนวนอนเป็นปี แนวตั้งเป็นเดือน ค่าในแต่ละจุดเป็น จำนวนผู้เสียชีวิตในเดือนกับปีนั้นๆ (3 คะแนน)

9. พิจารณาข้อมูลยอดเหตุการณ์อุบัติเหตุยานพาหนะ **accidents.xlsx** ที่ให้มา และเข้าถึงเว็บ https://en.wikipedia.org/wiki/Provinces_of_Thailand เพื่อใช้สำหรับแปลงชื่อจังหวัดภาษาไทยให้เป็นอังกฤษ และนำพื้นที่ (Area) และจำนวนประชากร (ปี 2022) มาใช้งานร่วมกับข้อมูลอุบัติเหตุ

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	id	Age	Sex	Nationality	lc	DeadDate	DateRec	SubDistrict	District	Province	AccLat	Acclong	ICD-10	Vehicle
2	8648189	0	1		99	30/4/2013				กระบี่			V244	รถจักรยานยนต์
3	8692138	0	1			26/10/2015	26/10/2015	อ่าวลึกใต้	อ่าวลึก	กระบี่	8.41986	8.419858		ไม่ระบุพาหนะ
4	8713884	0	1			28/10/2016	28/10/2016	โคกยาง	เหนือคลอง	กระบี่	8.08073	8.08073		ไม่ระบุพาหนะ
5	9618929	0				27/7/2018	27/7/2018	เขาพนม	เขาพนม	กระบี่	8.24014	8.240136		รถจักรยานยนต์
6	9800090	0	2		99	28/9/2017	28/9/2017	ปากสัถ	เหนือคลอง	กระบี่	8.0423	8.042296	V499	รถยนต์
7	8648889	0	1			11/5/2013	11/5/2013	มินบุรี	มินบุรี	กรุงเทพมหานคร	13.7317	13.73173		ไม่ระบุพาหนะ
8	8648891	0	2			13/4/2013	13/4/2013	กระทุ่มราย	หนองจอก	กรุงเทพมหานคร	13.8225	13.82246		ไม่ระบุพาหนะ
9	8648899	0	1			5/6/2013	5/6/2013	มินบุรี	มินบุรี	กรุงเทพมหานคร	13.8184	13.81839		ไม่ระบุพาหนะ
10	8648928	0	1			30/10/2013	30/10/2013	จิมพลี	ดลิ่งชัน	กรุงเทพมหานคร	13.7812	13.78121		ไม่ระบุพาหนะ
11	8649045	0	1			16/11/2013	16/11/2013	สามเสนใน	พญาไท	กรุงเทพมหานคร	13.7901	13.79009		ไม่ระบุพาหนะ
12	8649081	0	1			20/8/2013	20/8/2013	สามวาตะวันตก	คลองสามวา	กรุงเทพมหานคร	13.857	13.857		ไม่ระบุพาหนะ
13	8649085	0	1			16/7/2013	16/7/2013	สีกัน	ดอนเมือง	กรุงเทพมหานคร	13.9447	13.94465		ไม่ระบุพาหนะ
14	8649154	0	1			6/5/2013	6/5/2013	ออเงิน	สายไหม	กรุงเทพมหานคร	13.8872	13.8872		ไม่ระบุพาหนะ
15	8649265	0	1			31/5/2013	31/5/2013	สามเสนใน	พญาไท	กรุงเทพมหานคร	13.7908	13.79077		ไม่ระบุพาหนะ
16	8649282	0	1			6/11/2013	6/11/2013	ดินแดง	ดินแดง	กรุงเทพมหานคร	13.7646	13.7646		ไม่ระบุพาหนะ
17	8649290	0	1			14/12/2013	13/12/2013	วัดโสมนัส	ป้อมปราบศัตรูพ่าย	กรุงเทพมหานคร	13.7421	13.74212		ไม่ระบุพาหนะ
18	8649295	0	1			15/6/2013	15/6/2013	บางแค	บางแค	กรุงเทพมหานคร	13.6998	13.69978		ไม่ระบุพาหนะ
19	8649305	0	2			7/8/2013	7/8/2013	บางบอน	บางบอน	กรุงเทพมหานคร	13.6614	13.66144		ไม่ระบุพาหนะ
20	8649624	0	1			23/9/2013	23/9/2013	ดอนเมือง	ดอนเมือง	กรุงเทพมหานคร	13.9221	13.92211		ไม่ระบุพาหนะ
21	8649677	0	2			1/7/2013	30/6/2013	บางไผ่	บางแค	กรุงเทพมหานคร	13.7505	13.75046		ไม่ระบุพาหนะ
22	8649982	0	2			24/10/2013	24/10/2013	คลองเตย	คลองเตย	กรุงเทพมหานคร	13.9313	13.93134		ไม่ระบุพาหนะ

ตัวอย่างข้อมูลของ accident.xlsx

Seal	Name	Name (in Thai)	Population (December 2022) ^[1]	Area (km ²) ^[2]	Population density	Namesake town/city	HS ^[6]	ISO ^[7]	FIPS
	Bangkok (special administrative area)	กรุงเทพมหานคร	5,702,000	1,564	3,623	Bangkok	BKK	TH-10	TH40
	Amnat Charoen	อำนาจเจริญ	402,000	3,290	115	Amnat Charoen	ACR	TH-37	TH77
	Ang Thong	อ่างทอง	301,000	950	294	Ang Thong	ATG	TH-15	TH35
	Bueng Kan	บึงกาฬ	450,000	4,003	106	Bueng Kan	BKN	TH-38	TH81
	Buriram	บุรีรัมย์	1,623,000	10,080	159	Buriram	BRM	TH-31	TH28
	Chachoengsao	ฉะเชิงเทรา	754,000	5,169	139	Chachoengsao	CCO	TH-24	TH44

ตัวอย่างข้อมูลของ https://en.wikipedia.org/wiki/Provinces_of_Thailand

- 9.1. จำนวนครั้งการเกิดอุบัติเหตุในเขตกรุงเทพมหานคร โดยเรียงลำดับตามจำนวนครั้งการเกิดสำหรับประเภท ยานพาหนะ (Vehicle) จากมากไปน้อย โดยนำเสนอเป็นทั้งตัวเลข และ กราฟที่เหมาะสม (3 + 2 = 5 คะแนน)
- 9.2. นำเสนอตารางสรุปข้อมูลจำนวนประชากร พื้นที่ ความหนาแน่น จำนวนครั้งอุบัติเหตุ จำนวนครั้งอุบัติเหตุต่อจำนวนประชากรหนึ่งแสนคน และ จำนวนครั้งอุบัติเหตุต่อพื้นที่ ของแต่ละจังหวัด โดยเรียงลำดับตามคอลัมน์ จำนวนครั้งอุบัติเหตุต่อจำนวนประชากรหนึ่งแสนคน (Acct/Pop(100000 คน)) ดังตัวอย่างต่อไปนี้ (6 คะแนน)

Province	Population	Area	Density	No.accidents	Acct/Pop100000	Acct/Area
Rayong	727000	3666	201	4822	663.274	1.315
Nakhon Nayok	224000	2141	122	1195	533.482	0.558
Prachuap Khiri Khan	530000	6414	88	2752	519.245	0.429
Prachinburi	506000	5026	99	2614	516.601	0.520
Chonburi	1603000	4508	346	8226	513.163	1.825
Chanthaburi	572000	6415	84	2767	483.741	0.431
Saraburi	708000	3499	185	3414	482.203	0.976
Sing Buri	198000	817	255	926	467.677	1.133
Chachoengsao	754000	5169	139	3500	464.191	0.677
Phuket	387000	547	762	1741	449.871	3.183

9.3. จากข้อย่อยที่แล้ว (9.2) ท่านคิดว่าจำนวนครั้งอุบัติเหตุมีความสัมพันธ์กับข้อมูล {Population, Area, Density} ไດ
 มากน้อยกว่ากันอย่างไร (3 คะแนน)

----- สิ้นสุดคำถาม -----

ตัวอย่าง sales.csv

	A	B	C	D	E	F
1	date	date_block_num	shop_id	item_id	item_price	item_cnt_day
2	09.05.2013	4	51	17823	370	1
3	20.05.2013	4	27	7218	1599	1
4	23.04.2015	27	58	4808	399	1
5	25.04.2013	3	25	6624	1699	1
6	31.05.2014	16	58	1803	499	1
7	06.01.2014	12	54	11592	399	1

ตัวอย่างไฟล์ Shop.csv

	A	B
1	shop_name	shop_id
2	Daily Mart	0
3	All-Day Supplies Shop	1
4	Daily Life Market	2
5	Neat Necessities Store	3
6	Ready-to-Go Shop	4
7	Practical Picks Store	5

ตัวอย่างไฟล์ items.csv

	A	B	C
1	item_name	item_id	item_category_id
2		0	40
3		1	76
4		2	40
5		3	40
6		4	40
7		5	40

ตัวอย่างไฟล์ item_categories.csv

	A	B
1	item_category_name	item_category_id
2	Snacks	0
3	Beverages	1
4	Cereal and Breakfast	2
5	Pasta and Rice	3
6	Condiments and Sauces	4
7	Baking Supplies	5

ตัวอย่างไฟล์ gdp_per_capita_growth.csv

	A	B	C	D	E	F	G	H	I	J	K	L
1	country	1801	1802	1803	1804	1805	1806	1807	1808	1809	1810	1811
2	Aruba	0	0	0	0	0	0	0	0	0	0	0
3	Afghanistan	0	0	0	0	0	0	0	0	0	0	0
4	Angola	0.425	0.425	0.425	0.425	0.425	0.425	0.425	0.425	0.425	0.425	0.425
5	Anguilla	0	0	0	0	0	0	0	0	0	0	0
6	Albania	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104
7	Andorra	0.166	0.166	0.166	0.166	0.166	0.166	0.166	0.166	0.166	0.166	0.166
8	Netherlands Antill	0	0	0	0	0	0	0	0	0	0	0
9	UAE	0.0236	0.0236	0.0236	0.0236	0.0236	0.0236	0.0236	0.0236	0.0236	0.0236	0.0236
10	Argentina	0	0	0	0	0	0	0	0	0	0	0

ตัวอย่างไฟล์ covid19.csv

	A	B	C	D	DO	DP	DQ	DR	DS	DT
1	Province/State	Country/Region	Lat	Long	5/15/20	5/16/20	5/17/20	5/18/20	5/19/20	5/20/20
11	Australian Capital	Australia	-35.4735	149.0124	107	107	107	107	107	107
12	New South Wales	Australia	-33.8688	151.2093	3074	3075	3076	3078	3081	3082
13	Northern Territor	Australia	-12.4634	130.8456	29	29	29	29	29	29
14	Queensland	Australia	-27.4698	153.0251	1055	1055	1057	1057	1058	1058
15	South Australia	Australia	-34.9285	138.6007	439	439	439	439	439	439
16	Tasmania	Australia	-42.8821	147.3272	228	228	228	228	228	228
17	Victoria	Australia	-37.8136	144.9631	1551	1558	1564	1573	1573	1581

ตัวอย่างไฟล์ accidents.xlsx

	A	B	C	D	E	F	G	H	I	J	K	L	
1	id	Age	Sex	NationalityId	DeadDate	DateRec	SubDistrict	District	Province	AccLat	AccLong	ICD-10	Vehicle
2	8648189	0	1	99	30/4/2013				กระบี่			V244	รถจักรยานยนต์
3	8692138	0	1		26/10/2015	26/10/2015	อำเภอไทร	อำเภอ	กระบี่	8.41986	8.419858		ไม่ระบุพาหนะ
4	8713884	0	1		28/10/2016	28/10/2016	โคกยาง	เหนือคลอง	กระบี่	8.08073	8.08073		ไม่ระบุพาหนะ
5	9618929	0			27/7/2018	27/7/2018	เขาพนม	เขาพนม	กระบี่	8.24014	8.240136		รถจักรยานยนต์
6	9800090	0	2	99	28/9/2017	28/9/2017	ปากสั	เหนือคลอง	กระบี่	8.0423	8.042296	V499	รถยนต์
7	8648889	0	1		11/5/2013	11/5/2013	มีนบุรี	มีนบุรี	กรุงเทพมหานคร	13.7317	13.73173		ไม่ระบุพาหนะ