

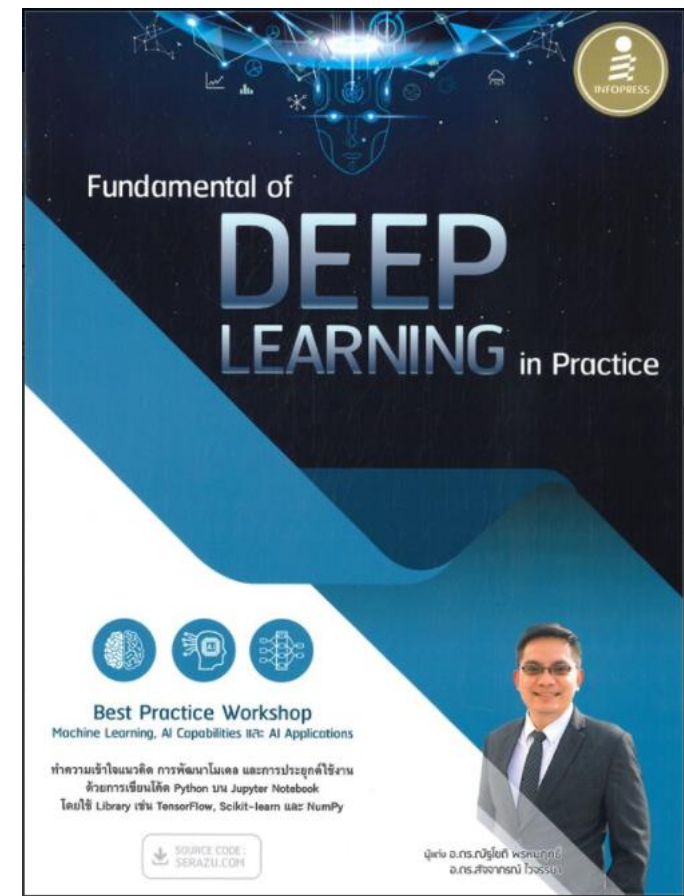
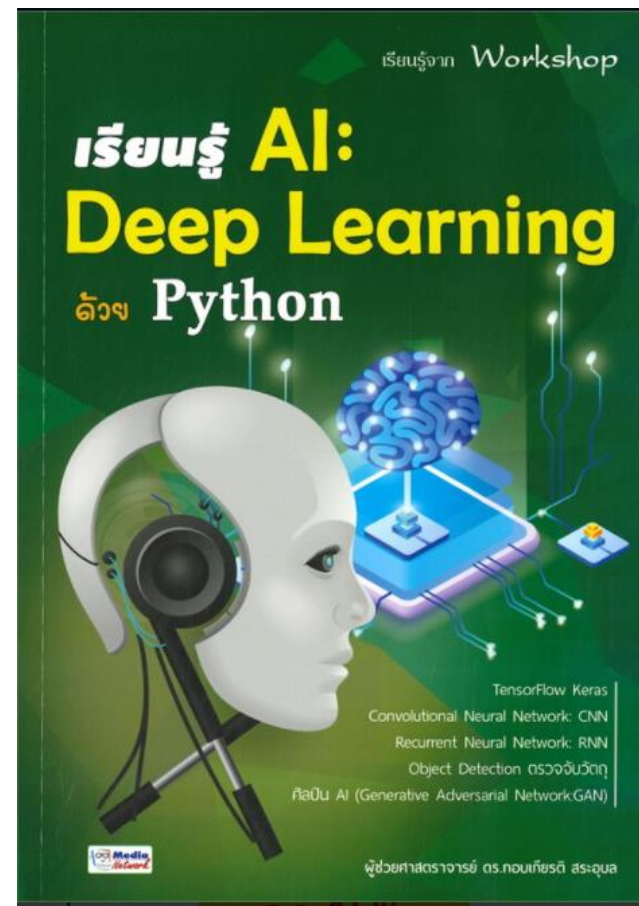
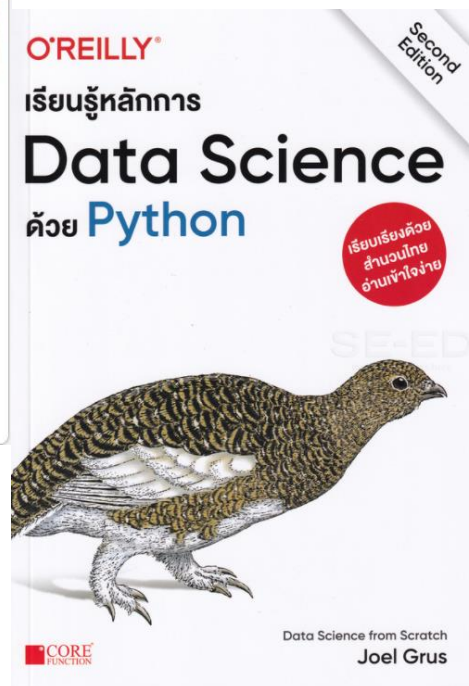
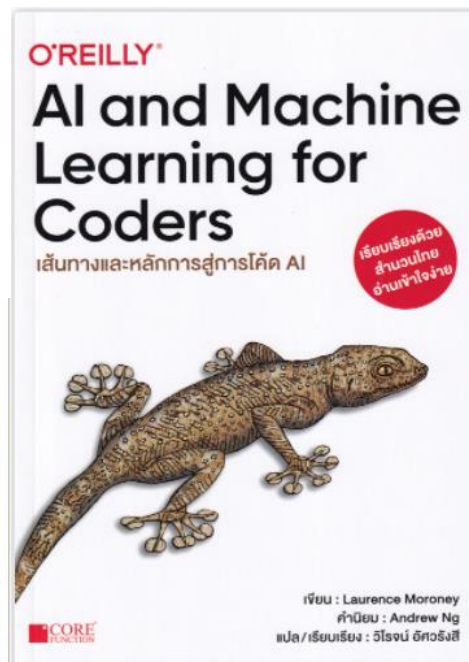
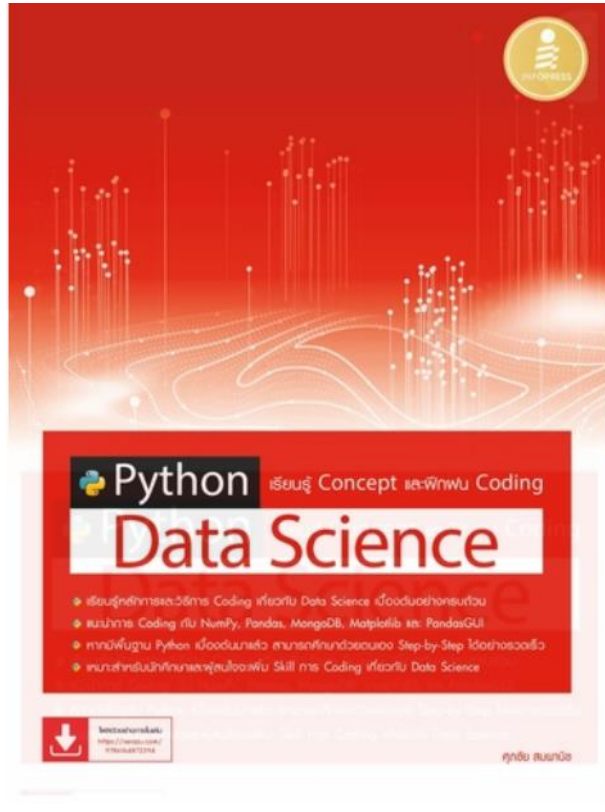
LECTURE 7

DATA EXPLORATION, VISUALIZATION AND VISUAL ANALYSIS

DR. PRAPASSORN TANTIPHANWADI

INDUSTRIAL ENGINEERING, FACULTY OF ENGINEERING AT KHAMPAENGSSEN

DECEMBER 2565



CONTENT

- พื้นฐานการทำงานข้อมูลด้วย Pandas (บท3)
- DataFrame ของ Pandas (บท4)
- การทำงานกับไฟล์ Excel (บท5)
- ทำงานข้อมูลในไฟล์ CSV (บท6)
- การทำงานกับข้อมูลสูญหายและค่าผิดปกติ (บท9)
- พื้นฐานการแสดงผลแบบกราฟ (บท14)

พื้นฐานการทำงานข้อมูลด้วย PANDAS

- Pandas เป็น library พื้นฐาน ถูกสร้างเพื่อมาทำงานกับข้อมูล
 - ด้านการจัดการข้อมูลจากแหล่งที่หลากหลาย
 - งานด้านการคำนวณ
 - การวิเคราะห์ทางสถิติ
- การติดตั้ง
 - Pip install pandas
- ฟังก์ชัน DataFrame()
 - Example 3-1การแสดงผลข้อมูลจาก array 2 มิติ (หน้า 56)

พื้นฐานการทำงานข้อมูลด้วย PANDAS

- ฟังก์ชัน DataFrame()
 - การถอด column หรือ row ออกจาก DataFrame ด้วยฟังก์ชัน drop()
 - ตัวอย่าง ถอด column และ row (หน้า 57-59)
 - การถอดค่าซ้ำออกจาก DataFrame ด้วย drop_duplicates() (หน้า 60-61)
 - การเปลี่ยนชื่อ column ใน DataFrame ด้วย rename() (หน้า 61-62)
 - การสร้างฟังก์ชันการทำงานกับข้อมูลโดยอาศัยฟังก์ชัน apply() (ตัวอย่าง 3-3 หน้า 62-63)
- การสร้าง lambda ทำงานกับข้อมูลด้วย apply() (หน้า 64)
- การทำงานกับข้อมูลตามแกน X และ แกน Y (หน้า 65)
- การทำงานกับข้อมูลตามแนว column ด้วย applymap() (หน้า 66)
- โครงสร้างข้อมูลแบบ Series ของ Pandas (หน้า 67-69)
 - Series แบบกำหนดค่า Index เอง (หน้า 70)
 - การสร้าง Series จากโครงสร้างข้อมูล Dictionary (หน้า 71)
 - การอ่านข้อมูลจาก series (หน้า 72-77)

DATAFRAME ของ PANDAS

- DataFrame เป็นโครงสร้างข้อมูลที่มีลักษณะเหมือนตาราง
 - พื้นฐานการสร้าง DataFrame ของ Pandas (หน้า 79)
 - การสร้าง DataFrame แบบหลาย Column (หน้า 81)
 - การสร้าง DataFrame จากรายการ list แบบหลาย column (หน้า 83-86)
 - การตรวจสอบข้อมูลเบื้องต้นของ DataFrame (หน้า 86-89)
 - การตรวจสอบชนิดข้อมูลของ DataFrame ด้วยคุณสมบัติ dtypes (หน้า 89-92)
 - การสร้าง DataFrame จาก Dict (หน้า 92-95)
 - การรวม DataFrame ด้วย append() (หน้า 95-97)
 - การรวม DataFrame ด้วย concat() (หน้า 98-100)

- พื้นฐานการอ่านข้อมูลจาก DataFrame ด้วย loc
 - loc ทำหน้าที่อ่านข้อมูล มีคุณสมบัติ
 - กรณีอ่านข้อมูลแถวใดแถวหนึ่ง (หน้า 100)
 - ต้องการข้อมูลหลายแถว (หน้า 101)
 - ระบุมุม และระบุมุมกลับ
 - ระบุมุม
 - ระบุมุมและระบุมุมกลับ
 - ระบุมุมด้วยค่าตรรกะ (หน้า 104)
 - ระบุมุมแบบมีเงื่อนไข (หน้า 105)

DATAFRAME ของ PANDAS

- DataFrame เป็นโครงสร้างข้อมูลที่มีลักษณะเหมือนตาราง
 - พื้นฐานการจัดกลุ่มข้อมูลใน DataFrame ด้วย `groupby()` (หน้า 105)

การทำงานกับไฟล์ EXCEL

- การอ่านข้อมูลจากไฟล์ Excel ด้วย xlrd
 - pip install xlrd (หน้า109-113) Product.xlsx
 - การเลือกข้อมูลเป็นช่วงจากไฟล์ excel (หน้า 114-117)
 - การค้นหาข้อมูลด้วย match(), contain() ของ Pandas (หน้า 118-120)
- การสร้างไฟล์ Excel จาก DataFrame ด้วย xlsxwriter
 - pip install xlsxwriter (หน้า121-113)
- การ export ข้อมูล จาก DataFrame ไปสู่โครงสร้างข้อมูลแบบอื่นๆ
 - โครงสร้างแบบ CSV (หน้า123) ตัวอย่าง 5-4 หน้า 128
- การ สร้างไฟล์ Excel และเพิ่มข้อมูลด้วย openpyxl
 - pip install openpyxl (หน้า 129-131)
- การ ใส่ข้อมูลแบบมีเงื่อนไข (หน้า 132)

การทำงานกับข้อมูลในไฟล์ CSV

- การอ่านข้อมูลจากไฟล์ CSV ด้วย read_csv() ของ Pandas Product.csv (หน้า 136)
 - การระบุเครื่องหมายในไฟล์ csv (หน้า 137)
 - การเลือกบาง column หรือบาง row (หน้า 138-142)
 - การเปลี่ยนชนิดข้อมูล column (หน้า 143)
 - การเปลี่ยนเป็นชนิดข้อมูลจัดกลุ่ม category ของ Pandas (หน้า 146)
- การเลือกข้อมูลจากไฟล์ CSV ด้วย คุณสมบัติ iloc (หน้า 148-151)
- ชนิดข้อมูลพื้นฐานของ python กับ Pandas

ชนิดข้อมูลของไพธอน	ชนิดข้อมูลของ Pandas	รายละเอียด
int	int64	เลขจำนวนเต็ม
float	float64	ตัวเลขทศนิยม
str	object	ข้อความ
bool	bool	ค่าทางตรรกะ

การทำงานกับข้อมูลสูญหายและค่าผิดปกติ

- ปัญหาของข้อมูลเสียหาย (หน้า 197)
 - การถอดข้อมูลสูญหายออกด้วย dropna (หน้า 198)
 - การถอดข้อมูลสูญหาย แบบใช้ตัวแปรเติม (หน้า 199)
 - การยินยอมให้มีค่าสูญหาย (หน้า 200)
 - การถอดค่าแบบกำหนดแกน X หรือแกน Y (หน้า 201)
 - การคำนวณปริมาณข้อมูลสูญหาย (หน้า 202-203)
 - การตรวจสอบข้อมูลสูญหายด้วย isnull() (หน้า 204)
 - การเติมค่าที่สูญหายด้วย interpolate() (หน้า 205)
 - การเติมค่าที่สูญหายด้วย fillna() (หน้า 206)
- ทำความรู้จักกับข้อมูลที่มีค่าผิดปกติ Outlier (หน้า 210)
- การคำนวณค่าเฉลี่ยเลขคณิตแบบตัดค่าความเสียหายด้วย nanmean() (หน้า 212)

พื้นฐานการแสดงผลแบบกราฟด้วย MATPLOTLIB

- pip installmatplotlib (หน้า 291)
 - Scatter plot (หน้า 292)
 - กราฟจากข้อมูลหลายชุด (หน้า 294)
 - การกำหนดขนาดจุด (หน้า 297)
 - การกำหนดความเข้มจางของข้อมูล (หน้า 298)
 - Bar graph (หน้า 299)
 - Pie graph (หน้า 302)
 - Line graph (หน้า 304)
 - การกำหนดข้อมูล 1 แกน (หน้า 308)
 - การสร้างกราฟเส้นและ mark ข้อมูล (หน้า 309)
 - การปรับแต่งกราฟเป็นเส้นประ (หน้า 310)
 - การปรับแต่งรูปแบบกราฟด้วย linestyle (หน้า 313)
 - กราฟเส้นของข้อมูลหลายชุด (หน้า 315)
 - การแยกกราฟเส้นของข้อมูลแต่ละชุด(หน้า 316)
 - การสร้าง Boxplot

HOMEWORK

➤ จงสร้างกราฟจากข้อมูลเหล่านี้ในไฟล์ excel หรือ csv

1. Scatter diagram Quality & pH
2. Matrix diagram of all parameters

Quality	pH	Total SO ₂	Color Density	Color
19.2	3.85	66	9.35	5.65
18.3	3.75	79	11.15	6.95
17.1	3.88	73	9.40	5.75
15.2	3.66	86	6.40	4.00
14.0	3.47	178	3.60	2.25
13.8	3.75	108	5.80	3.20
12.8	3.92	96	5.00	2.70
17.3	3.97	59	10.25	6.10
16.3	3.76	22	8.20	5.00
16.0	3.98	58	10.15	6.00
15.7	3.75	120	8.80	5.50
15.3	3.77	144	5.60	3.35
14.3	3.76	100	5.55	3.25
14.0	3.76	104	8.70	5.10
13.8	3.90	67	7.41	4.40
12.5	3.80	89	5.35	3.15
11.5	3.65	192	6.35	3.90
14.2	3.60	301	4.25	2.40
17.3	3.86	99	12.85	7.70
15.8	3.93	66	4.90	2.75

HOMEWORK

- จงสร้างกราฟจากข้อมูลเหล่านี้ในไฟล์ excel หรือ csv

3. Histogram

Layer Thickness (Å) on Semiconductor Wafers

438	450	487	451	452	441	444	461	432	471
413	450	430	437	465	444	471	453	431	458
444	450	446	444	466	458	471	452	455	445
468	459	450	453	473	454	458	438	447	463
445	466	456	434	471	437	459	445	454	423
472	470	433	454	464	443	449	435	435	451
474	457	455	448	478	465	462	454	425	440
454	441	459	435	446	435	460	428	449	442
455	450	423	432	459	444	445	454	449	441
449	445	455	441	464	457	437	434	452	439

HOMework

➤ จงสร้างกราฟจากข้อมูลเหล่านี้ในไฟล์ excel หรือ csv

4. Boxplot of 3 group of pizza data

Old Chicago	Calories	Old Chicago	Calories	Black-eye Pea	Calories
Best Chicken Wings	1562	BBQ Chicken <u>delite</u>	320	½ Pound Burger	961
Bacon Cheese Burger	1241	Beef Taco Grande	636	8 oz. <u>Sirioin</u> Steak	290
Caesar Salad	835	Big Murphy Stuffed	760	Buffalo Chicken	337
Cheese Garlic Bread	1063	Cheese <u>Delite</u> Thin	260	Caesar Salad	442
Chicken Blat Wrap	651	Chicago – Style Stuffed	740	<u>Charboiled</u> Steak	771
Chicken <u>Rustica</u>	513	Combination Calzone	900	Chicken <u>Pitpie</u>	656
Applewood BBQ	1481	<u>Gourmet</u> Chicken	640	Fried Catfish	496
Chicago 7	1341	<u>Gourmet</u> Italian	720	Grilled Chicken-rice	788
Chicken Artichoke	1436	<u>Gourmet</u> Veggie	600	Grilled <u>Salomn</u> -rice	763
Double <u>Deckeroni</u>	1566	Lasagna	510	Meatloaf Dinner	420
Hawaiian	1298	Papa's All Meat	740	Casserole	735
Italian Grinder	1457				
Meat Me	1550				
Chicago 7 Turkey	859				
Bacon Club	644				
Alfredo	1251				
Classic Meat	1052				
Spaghetti Meatball	1084				
BB Bacon Cheddar	842				
Burger	751				
The Classic Calzone	1215				

HOMEWORK

➤ จงสร้างกราฟจากข้อมูลเหล่านี้ในไฟล์ excel หรือ csv

5. Time series

Year	Quarter	Sales amount
1	2	8,000
1	3	13,000
1	4	23,000
2	1	34,000
2	2	15,000
2	3	23,000
2	4	28,000

Year	Quarter	Sales amount
3	1	43,000
3	2	22,000
3	3	23,000
3	4	42,000
4	1	51,000

HOMEWORK

➤ จงสร้างกราฟจากข้อมูลเหล่านี้ในไฟล์ excel หรือ csv

6. Pareto (all items)

7. Pie chart with separated segment
(grouping to 5 items)

Complaints	Number of items
Heating	10
Reservations	155
Décor	2
Room Service	287
TV set	23
Internet	17
Bed linen	16
Pillows	5
Towels	4
Furniture	9
Telephone	6
Cleaning	111
Noise	42
Documentation	1
Other	3

THE END