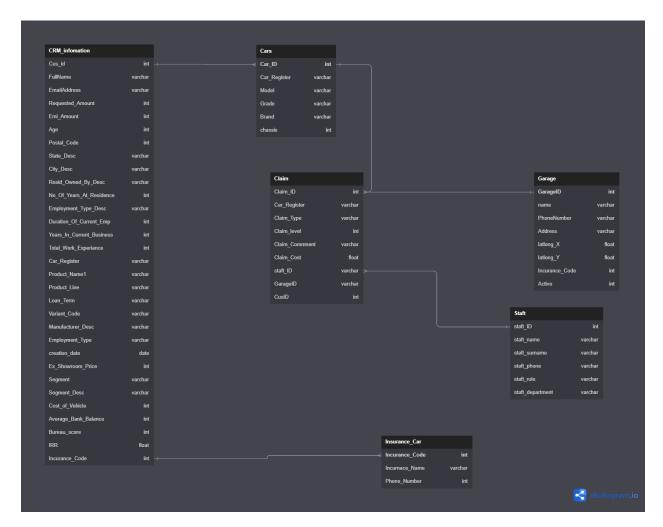
Problem

ปัจจัยในการนำเสนองายรถยนต์ของธนาคาร The one bank เป็นทำธุรกิจด้านการเงินและleasing ใช้ในแผนก callcenter

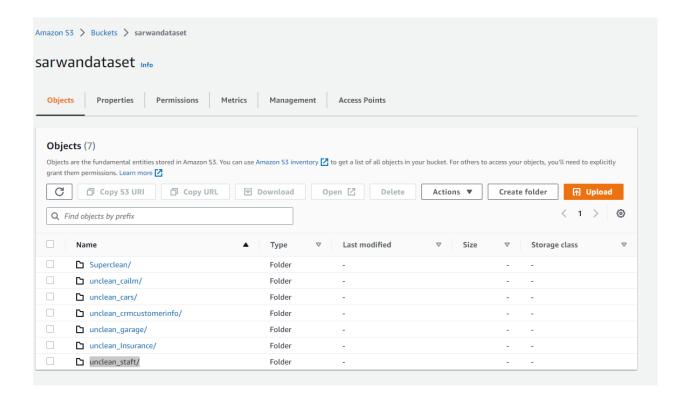
Data model



ขั้นตอนการทำงานดังนี้

Step1

- 1. Data source น้ำ Raw data วางไว้ที่ s3 เพื่อเป็น data source จำลองในการดึงข้อมูลมาใช้งาน และสร้าง Bucket ใน s3
- 2. ทำการสร้าง Folder cleaned เพื่อเก็บ ข้อมูลดิบ

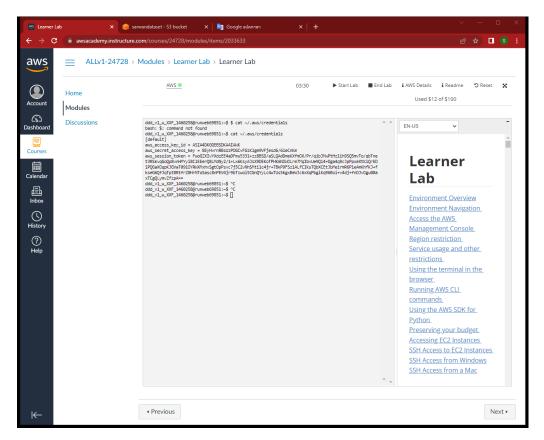


Step2

3.Dataleake Zone

เป็นส่วนที่ใช้จัดเก็บไฟล์ต่างๆที่ได้ดึงข้อมูลมาและมาเก็บไว้เพื่อจะไปทำ ETL ต่อ โดยจะทำการสร้าง datalake_s3.py เพื่อดึงข้อมูลจาก Folder landingZone มาจัดเก็บไว้ใน s3 โดยใช้ Spark ในการดึงข้อมูลและสร้างตาราง

- -ทำการสร้าง Folder superclean เพื่อเก็บ calean data
- -ตั้งค่าการเชื่อมต่อ AWS โดยใช้คำสั่ง \$ cat ~/.aws/credentials และนำkey ดังนี้ไปใส่ใน code python
- -aws_access_key_id
- aws_secret_access_key
- aws_session_token



ทำ การรันผ่าน docker compose และเข้าไปรันไฟล์ที่ port 8888

cd capstone-project

python -m venv ENV

source ENV/bin/activate

pip install -r requirements.txt

```
(ENV) gitpod /workspace/swu-ds525/Capstone_FinalProject (main) $ pip install -r prerequisite/requirements.txt
Collecting numpy==1.23.2
Using cached numpy-1.23.2-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (17.1 MB)
Collecting psycopg2==2.9.3
Using cached psycopg2-2.9.3.tar.gz (380 kB)
Preparing metadata (setup.py) ... done
Collecting psycopg2-binary==2.9.5
```

docker-compose up

port 8888

สร้างไฟล์datalake_s3.py บน Jupyter

```
[2]: from pyspark import SparkConf
            from pyspark.sql import SparkSession
[3]: # cat ~/.aws/credentials
            aws_access_key_id = 'ASIA4DKGGEBSDKAAIAWK'
            aws_secret_access_key = '8EyHvrnRBsUzPOGCvFGzKlgm9VFjesd6/6leCnKm'
aws_session_token = 'FwoGZXIvYXdzEE4aDFmu5331+zz8BSD/aSLQAdDmeKXfmOX/Pr/qibJMwPtHz1ih9SQ5mvTo/qbTnet3RSkvqGoQp8vn
            4
[7]: conf = SparkConf()
            conf.set("spark.jars.packages", "org.apache.hadoop:hadoop-aws:3.2.2")
            conf.set ("spark.hadoop.fs.s3a.aws.credentials.provider", "org.apache.hadoop.fs.s3a.TemporaryAWSCredentialsProvider", "org.apache.hadoop.fs.saa.TemporaryAWSCredentialsProvider", "org.a
            conf.set("spark.hadoop.fs.s3a.access.key", aws_access_key_id)
            conf.set("spark.hadoop.fs.s3a.secret.key", aws_secret_access_key)
            conf.set("spark.hadoop.fs.s3a.session.token", aws_session_token)
[7]: <pyspark.conf.SparkConf at 0x7f2c3c409660>
[8]: bucket = "s3a://sarwandataset"
            landing_zone_customer = f"{bucket}/unclean_crmcustomerinfo/"
            exportData = f"{bucket}/Superclean/"
[9]: spark = SparkSession.builder.config(conf=conf).getOrCreate()
10]: #data = spark.read.option("header", "true").option("multiline", "true").csv(cleaned_zone)
            #data.createOrReplaceTempView("staging_data")
            data_customer = spark.read.option("header","true").option("sep",",").csv(landing_zone_customer)
            data_customer.createOrReplaceTempView("staging_customer")
             landing zone cailm = "s3a://sarwandataset/unclean cailm/Cailm.csv"
12]: spark = SparkSession.builder.config(conf=conf).getOrCreate()
13]: data_cailm = spark.read.option("header","true").option("sep",",").csv(landing_zone_cailm)
            data_cailm.createOrReplaceTempView("stagging_cailm")
```

```
[16]: landing_zone_cars = "s3a://sarwandataset/unclean_cars/Cars.csv"
      landing_garage = "s3a://sarwandataset/unclean_garage/Garage.csv"
      landing_zone_Insurance = "s3a://sarwandataset/unclean_Insurance/Insurance_car.csv"
      landing_zone_staft = "s3a://sarwandataset/unclean_staft/Staft.csv/"
[17]: data_cars = spark.read.option("header","true").option("sep",",").csv(landing_zone_cars)
      data_cars.createOrReplaceTempView("stagging_cars")
[18]: data_garage = spark.read.option("header","true").option("sep",",").csv(landing_garage)
      data_garage.createOrReplaceTempView("stagging_garage")
[19]: data_incurance = spark.read.option("header","true").option("sep",",").csv(landing_zone_Insurance)
      data_incurance.createOrReplaceTempView("stagging_Incurance")
[20]: data_staft = spark.read.option("header","true").option("sep",",").csv(landing_zone_staft )
      data_staft.createOrReplaceTempView("stagging_staft")
[21]: data_customer.printSchema()
      data_cailm.printSchema()
      data cars.printSchema()
      data_garage.printSchema()
      data_incurance.printSchema()
      data staft.printSchema()
 |-- Cus_ID: string (nullable = true)
 |-- FullName: string (nullable = true)
 |-- EmailAddress: string (nullable = true)
 |-- Requested_Amount: string (nullable = true)
 |-- Emi Amount: string (nullable = true)
 |-- Age: string (nullable = true)
 -- Postal_Code: string (nullable = true)
 |-- Applicant_State_Desc: string (nullable = true)
 |-- Applicant_City_Desc: string (nullable = true)
 |-- Resid_Owned_By_Desc: string (nullable = true)
 |-- No_Of_Years_At_Residence: string (nullable = true)
 -- Employment_Type_Desc11: string (nullable = true)
 |-- Duration_Of_Current_Emp: string (nullable = true)
 |-- Years_In_Current_Business: string (nullable = true)
 |-- Total_Work_Experience: string (nullable = true)
 |-- Car_Register: string (nullable = true)
 |-- Product Name1: string (nullable = true)
 |-- Product_Line: string (nullable = true)
 |-- Loan_Term: string (nullable = true)
 |-- Variant Code: string (nullable = true)
 |-- Manufacturer_Desc: string (nullable = true)
 -- Employment_Type_Desc21: string (nullable = true)
 |-- application_creation_date: string (nullable = true)
 |-- Ex_Showroom_Price: string (nullable = true)
 |-- Segment: string (nullable = true)
 |-- Segment_Desc: string (nullable = true)
 -- Cost Of Vehicle: string (nullable = true)
 |-- Average_Bank_Balance: string (nullable = true)
 |-- Bureau_score: string (nullable = true)
 -- IRR: string (nullable = true)
```

|-- insurance_Code: string (nullable = true)

```
root
 |-- Claim_ID: string (nullable = true)
 |-- Car_Register: string (nullable = true)
 -- Claim_Type: string (nullable = true)
 |-- Claim_level: string (nullable = true)
 |-- Claim_Comment: string (nullable = true)
 |-- Calim_Cost: string (nullable = true)
 |-- GarageID: string (nullable = true)
 |-- CusID: string (nullable = true)
root
 |-- Car_ID: string (nullable = true)
 -- Car Register: string (nullable = true)
 |-- model: string (nullable = true)
 |-- Grade: string (nullable = true)
 |-- Brand: string (nullable = true)
 |-- chassis: string (nullable = true)
 |-- GarageID: string (nullable = true)
 |-- name: string (nullable = true)
 |-- PhoneNumber: string (nullable = true)
 |-- Address: string (nullable = true)
 |-- latlong_X: string (nullable = true)
 |-- latlong_Y: string (nullable = true)
 |-- IncuranceCode: string (nullable = true)
 |-- Active: string (nullable = true)
root
 |-- IncuranceCode: string (nullable = true)
 |-- IncuranceName: string (nullable = true)
 |-- Phonenumber: string (nullable = true)
root
 |-- Staff_ID: string (nullable = true)
 |-- Staff_Name: string (nullable = true)
 |-- Staff_Sername: string (nullable = true)
  |-- Staff_Phone: string (nullable = true)
  |-- Staff_Email: string (nullable = true)
  |-- Staff_Rule: string (nullable = true)
 |-- Staft_department: string (nullable = true)
```

D D > = 0	→ Code ∨																		Python 3 (ipyk
data customer.show(
data_costomer.snow()																			
data_cars.show()																			
data garage, show()																			
data_incurance.show	3																		
date staft.show()																			
	*																		
Cus_ID FullName	EmailAddress Request															_Business Total_No	/rk_Experience Ca	r_Register	Product_!
Product_Line Loan_Te							e Ex_Showroom_Price												
	*																		
	James1@gmail.com	200000	5750 40.			Thonburil	Bnegkok		SELFI		11/	SALARIED		*1		nulli	41	AND REES!	CHEVROLET
CHEVROLET SPARK L					SALARIED		4/17/2015	null		Compact	237342		ull] 59	4 24.7	101401	HOTEL	21	Her 0002]	CHEVROLET
	narlesl@email.com	1900000	null 41.2		SACHATO	bangkagil	Bnagkok	11011	SELF		11	RETIRED	OLL ST	null	707467	null	mul1	nua 86881	FORTUNER
TOYOTA - FORTUNES			TOYOT		RETIRED		4/23/2014	null		suvi	2400000		ull 75	4 24.7	100602	HOTT	MATT	Harr 8000]	PURTUNER
	secretiumail.com	16000001	null 39.		HELTHOU		4/23/2014] Bnarkok	MULL	SELFI			RETIRED	1111		199905	m-227	nulli	nu6532	
TOYOTA - FORTUNER			TOYOT		present l	bangkapi		TA CASE			2200000		and the same	null	100602	null	mull	nue532]	
	Frankißgmail.com	3000001	9050 3		RETIRED		4/25/2014	null	SELF	SUV	71	SALARIED	ull 27	9 16.5	100002]	null	161	nu42651	NISSAN
	48 NISSAN MICRA XLI		ISSAN	SALARIE		bangkapi 1/2/201	8nagkok 31 365756	I A2		365750		null	818 18.1			null	101	1184205]	NESSAN
	48 NISSAN MICHA XL Osephi@gmail.com	250000]		5ALARIE 10240	ni	bangkapi	Bnagkok	AZ	Compact SELF	365/30		EMPLOYED NON	818 18.1			3 [mulli	na7067	XV
XATO E81 301	XYLO E8		SELF EMPL			1/12/2013		MW!	MINI	5487501	2 500		853 [17.69]	100602		2)	entil	H27067	
	Thomasifignail.com	150000		3 10240		bangkapil	Bnagkok	(ADV)	SELFI		nel erre	EMPLOYED NON	000[11.00]	null		61	mul1	ns:10060	FORD
	FORD FIGO 1.4 LXI			PLOYED NON		1/17/2013	8nagkok 329258	A21	Compacti	3292501	29 SELF	null	777 14.66	100602		61	unii	Har10066	FORD
	Henryl@gmail.com/	439999		5 10240		bangkapi	Bnagkok i	A21	SELFI		eal rrie	EMPLOYED PRO	///[14.00]	nulli		15	mulli	ns:12815	MARUTI
MARUTI SHIFT	36 SWIFT DIERE LDI			EMPLOYED PRO.	242	1/21/201		EL A3		482750		null	817 [13.86]	100602		13	HULL	MMTT019	THEO I I
	Robertl@gmail.com	405000		10240				1 2	SELF!	482739		EMPLOYED NON	01/[15:00]	mull		10	null	ns:11055	MARUTI
	M6 RITZLDI			EMPLOYED NON	1	bangkapi 1/23/2013	Bnagkok 461500	A21	Compact	4615001	al acci	nulli	789 [14.37]	100602		70	MUIII	Hall0331	F94013
	dward1@gmail.com	276000		10240	8.	bangkapi	Bnagkok	ALI	SELFI		10 551 5	EMPLOYED NON]	100/14:3//	mull		51	null	nw10777	HYURDAI
YUNDAI VERNA	36 VERNA CRDI SXI			F EMPLOYED NON.	SSD	1/31/28		(e) A3				nulli	828 114,54		53	21	100.2.2	riman / / /	Ti Company
	Harryl@gmail.com	234666		10240		bangkapi	Bnagkok		SELF			EMPLOYED NON	000/24134	null	-:	51	mul1	nsi32428]	ALT
MARUTI ALTO LXII	36 ALTO LXI			ELF EMPLOYED NO	0 1	2/12/		737	A21 Compa		737	nul1	837 14.		602]	-		100001	
	alterifignail.com	618888		16 10240		bangkapi	8nagkok		ARENTSI			EMPLOYED NON		null		15	null	na:13143	HYUNDAT
CYLHOAI VERNAL	36 HYUNDAI VERNA XXI			F EMPLOYED NOW.	2000	2/13/20			3 Mid-Size			null	738 114,03		91	(46)			
	orthuri@gmail.com	286668		7 10240		bangkapi	Bnagkok		null			EMPLOYED NON		null		nu11	nul1	na:15814	HYUNDAI
HYUNDAI I 10	36 I-19			EMPLOYED NON		2/16/201		oull		243256		null	-1 16.2	100602					91400
1 14 Fred	Fredifignail.com	300000		8 10240		bangkapi	Bnagkok	, marri	SELFI			EMPLOYED NON	21 20.21	null		301	nu11	nu5913	MARUTI S
MARUTI SHIFT!	36 SWIFT VDI			EMPLOYED NON	31	2/14/201		A2		345756		nu211	788 14.1	188681					
	lberti@gmail.com	2920001		7 10240		bangkagil	Bnaskok		SELFI			EMPLOYED NON		null		15	mul1	na6814	SKODA
	6 FABIA AMB 1.21		wonal sere	EMPLOYED NON	1	2/23/2013	356000	ANI	Mid-Size	356000		oull!	518 17.3	100602					

++	+	+			++	++	+
Claim_ID	Car_Register	Claim_Type	Claim_level	Claim_Comment	Calim_Cost	GarageID	CusID
++	+	+			+	++	+
1	กมง 8682	inside	3	engine	350,000	G30	1
2	กมง 8688	external	1	car bumper	50,000	G12	2
3	กม6532	All	2	null	300,000	G12	3
4	กม4265	All	2	null	300,000	G12	4
5	กม7067	All	2	null	300,000	G12	5
6	กม10060	All	2	null	300,000	G12	6
7	กม12815	All	2	null	300,000	G12	7
8	กม11055	All	2	null	300,000	G12	8
9	กม10777	All	2	null	300,000	G12	9
10	กม12420	All	2	null	300,000	G12	10
11	กม13143	All	2	null	300,000	G12	11
13	กม15814	All	2	null	300,000	G12	13
14	กม5913	All	2	null	300,000	G12	14
15	กม6814	All	2	null	300,000	G12	15
16	กม7752	All	2	null	300,000	G12	16
17	กม6747	All	2	null	300,000	G12	17
18	กฒ8688	All	. 2	null	. 300,000	G12	18
19	กฒ6532	All	. 2	null	. 300,000) G12	19
20	กฒ4265	All	. 2	null	. 300,000	G12	20
21	กฒ7067	All	. 2	null	. 300,000) G12	21
++	+				+	++	+

only showing top 20 rows

++	+				+
Car_ID	Car_Register	model	Grade	Brand	chassis
1	กมง 8682	CHEVROLET SPARK	LT 1.0 BS-IV OBDII	CHEVROLET INDIA LTD	00001
2	กมง 8688	FORTUNER 4 WD	FORTUNER 4 WD	TOYOTA	00002
3	กม6532	null	FORTUNER 3.0 L	TOYOTA	00003
4	กม4265	NISSAN MICRA	NISSAN MICRA XL	NISSAN	00004
5	กม7067	XYLO E8	XYLO E8	MAHINDRA	00005
6	กม10060	FORD FIGO	FORD FIGO 1.4 LXI	FORD INDIA	00006
7	กม12815	MARUTI SWIFT	SWIFT DZIRE LDI	MARUTI	00007
8	กม11055	MARUTI RITZ	RITZLDI	MARUTI	00008
9	กม10777	HYUNDAI VERNA	VERNA CRDI SX	HYUNDAI	00009
10	กม12420	ALTO LXI	ALTO LXI	MARUTI	00010
11	กม13143	HYUNDAI VERNA	HYUNDAI VERNA XXI	HYUNDAI	00011
13	กม15814	HYUNDAI I 10	I-10	HYUNDAI	00012
14	กม5913	MARUTI SWIFT	SWIFT VDI	MARUTI	00013
15	กม6814	SKODA FABIA	FABIA AMB 1.2	SKODA	00014
16	กม7752	HYUNDAI - ACCENT	ACCENT EXECUTIVE	HYUNDAI	00015
17	กม6747	CHEVROLET CRUZE	CHEVROLET CRUZE	CHEVROLET INDIA LTD	00016
18	กฒ8688	SKODA YETI	SKODA YETI	SKODA	00017
19	กฒ6532	MARUTI ALTO	ALTO LX	MARUTI	00018
20	กฒ4265	SKODA SUPERB	SKODA SUPERB 1.8	SKODA	00019
21	กฒ7067	FIAT PUNTO	PUNTO DYNAMIC DI	FIAT INDIA LTD	00020
++	+	+			+

only showing top 20 rows

++	+-		-++
GarageID name PhoneNumber	Address latlong_X	latlong_Y IncuranceCode	e Active
+	+-		-++
G1 หจก. ลัคกี่การาจ 992341800 ตรอกคานเรือ	15 Sa 663840.1723 15246	62.021 100205	0
G2 บริษัท ม. กมลชัยก 992341801 44/18 Ra	n Inthra 672984.9832 19	34057.04 100501	0
G3 บริษัท มงคล ออโต้ 992341802 78/4 Ram	Inthra R 675834.3944 153	2509.412 100501	0
G4 บริษัท มหาธนเซอร์ 992341873 31, 17 Ra	rm Intra 675562.872 153	2541.286 100501	0
G5 อู่เปียกการาจ 992341874 47, 49 Rarm	Intra 675169.5519 15335	57.139 100501	0
G6 บริษัท เจ.อาร์.คา 992341875 170 83/19	ช. 39 ถ 676236.1881 153	3052.726 100501	0
G7 รามอินทรา การาจ 992341876 15 ซอย ส	ทดปลาเค้า 674738.8423 153	1840.087 100501	0
G8 บริษัท ร่วมเจริญอ 992341877 2 ถนนพหลโด	บธิน 46/ 673511.4083 1535	269.428 100501	0
G9 บริษัท แคทลียา คา 992341878 3/118 ซอ	ยวัชรพล, 674954.6384 153	3164.734 100501	0
G10 บริษัท ถาวรมงคล จ 992341879 2361 Lat	Phrao Rd 676973.9212 15	22993.735 100601	0
G11 หจก. พรีเมียร์ คา 992341880 966 Ramkha	mhaeng 679469.0515 1522	230.389 100602	0
G12 บริษัท เจริญพงษ์ 992341807 1 Ramkham	aeng Rd 676920.6491 1521	678.504 100602	0
G13 บริษัท อู่เอส.เอ 992341701 45 Phueng เ	4i 13 A 675011.9911 15150	89.601 100901	0
G14 บริษัท ใพชยนต์กา 992341702 459 Punna	9 Withi 1 674748.305 1	100901 100901	0
G15 อู่บอดีเทค 992341703 37 Ramkham	haeng R 685692.2121 1525	915.846 101001	0
G16 บริษัท มิลเลียน 992341704 14/5 Ramkha	mhaeng 685926.282 15263	37.874 101001	0
G17 บริษัท เจริญพัฒนา 992341705 9 Soi Suw	inthawon 685985.6703 152	7471.905 101001	0
G18 บริษัท สุขุมวิท เ 992341706 109 Soi Ban	g Na-T 677817.5382 15120	22.415 101001	0
G19 หจก .ส. ธนาวรรณยนต์ 992341707 132 Sam		•	L 0
G20 บริษัท พี.ไอ.ซี 992341708 18 Thanon	Rom Kla 687757.0708 1527	7888.693 101001	0
+			-++

only showing top 20 rows

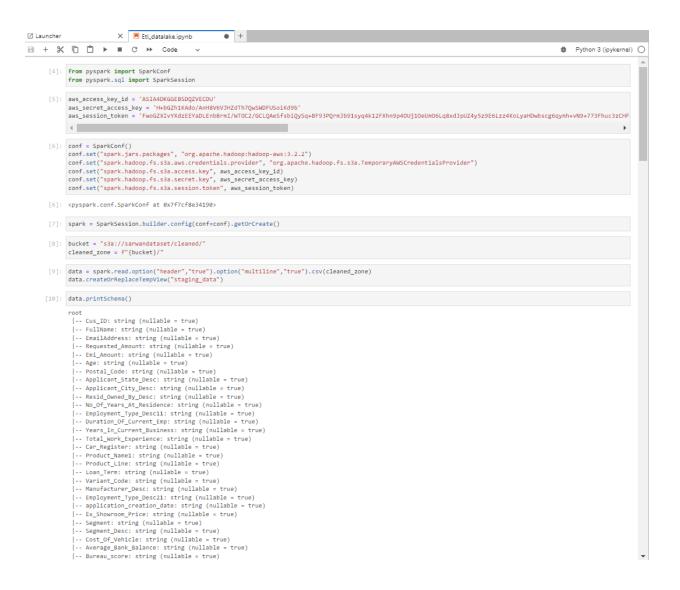
+		++		
IncuranceCode	IncuranceName	Phonenumber		
+	·	++		
100205	SOMPO	02-119-3000		
100501	Tokio marine	02-650-1407		
100601	Bangkok incurance	0 2285 8888		
100602	krungthai Exa	3 044 4000		
100901	viriyah incurance	\t+662-239-1558		
101001	Muangthai incurance	8 2274 9400		
101202	devez incurance	1291		
101401	MSIG	02 007 9009		
+	·			

+				+	+	+
St	aff ID St	taff Name S	staff Sername Staff Phone	Staff Email:	Staff Rule	Staft department
+						
1	5001	Alva	Adam 081-7222460	Adam@gmail.com	officer1	Callcenter
i.	5002	Harley	Black 081-7222461	Harley@gmail.com	Manager	Maintenance
i.	5003	Newton	Henri 081-7222462	Newton@gmail.com	officer2	Maintenance
i.	5004	Timothy	pink 081-7222463	Timothy@gmail.com	officer3	Maintenance
i.	5005	Marvin	white 081-7222464	Marvin@gmail.com	officer2	IT
i.	5006	Ross	Live 081-7222465	RoseLive@gmail.com	officer2	Marketing
i.	5007	Curtis	Orang 081-7222466	CTS@gmail.com	Manager	Callcenter
i.	5008	Edmund	Brazilliz 081-7222467	EdmondBra@gmail.com	Manager	Marketing
i.	5009	Jeff	Buffez 081-7222468	Jeff.B@gmail.com	officer1	Marketing
+						+

[23]: table_leagues - spark.sql("""		⑤ ↑ ↓ ☆ 〒 ■
+++		
[Cus_ID[Fallbase] EmailAddress[Requested_Amount[mid_Amount] Age [Postal_Code] Applicant_State_Desc[Applicant_City_Desc[Resid_Omod_By_Desc[No.of_Years_At_Besidence](Bub]aymont_Type_Descil[Invation_Of_Current_Emp[Years_In_Code] Product_Line[Loam_Term] Variant_Code] **BurdetCurrent_Desc[Invation_Of_Current_Emp[Years_In_Code]] **Dourson_Price[Segment_Segment_Desc[Cost_Of_Vahicle]Average_Bask_Balance Duranu_score)	Current_Business Total_Work_Experience	
1	null 5	nama 8682 CHEVROLET SPARK
2 Charles Charlesi@gmail.com	null null	naw 8688 FORTUNER 4 ND
3 George Georgel@gmsil.com 1600000 null 39.5 10240 bangkapi Bnagkok SELF null RETIRED null	null null	nu6532 null
4 Franki Franki Byssall.com 3000000 9050 35 10240 bangkapi Baagkok SEF 7 SALARIEO 16 NISSAN NICARA 48 NISSAN NICARA XI NISSAN NICARA XI	null 16	nu4265 NISSAN MICRA
5 Joseph Josephiggmail.com 250000 8750 34 10240 bengkapi Bnagiok 5ELF 2 SELF FUNCTO 10010 null XYO, ESI 36 XYO, ESI NHUTDAN, SELF FUNCTO NON 1/12/2031 548750 NVV NVV 549750 mill 83 17.09 100002	3 null	nu7067 XYLO E8
6 Thomas: Thomas:Edgesdi.com 1500000 5127 33 102340 bengkapi 8nagkok 5ELF 25 5ELF EMPLOYED NORTH null FOND FIELD 1861 FOND FIELD 1861 FOND FIELD NORTH 77714.66 1006002	6 null	nul0060 FORD FIGO
7 Henry Henryliges11.com 4300000 14552 45 10240 bangkap 8nagkok SEF 40 SEF DFU/DYD PRD null NABUT SER/FFT 36 SEF DFU/DYD PRD null 82713.86 100402	15 null	nu12815 MARUTI SWIFT
8 Sobert Sobert@gmeil.com 405900 13642 66 18230 bengkapi 8negkok 5ELF 9 SELF EPE/DYTE NOW null HABUT ENTZI 36 8TITLOI 8TITLO	10 null	nw11055 MARUTI RITZ
9 Edward[Eggmall.com 276000 9547 34 18240 bangkard 8nagkok SELF 30 SELF 9H_DVFE NDN null NHYBOAT SELF 9H_DVFE NDN 100602	5 null	nu10777 HYUNDAI VERNA
1.0 Herry Herry18gmall.com 234000 7997 34 10240 bangkapi 5547 655 554 EFF-0770 DRL 701 100000000000000000000000000000	5 null	nw12420 ALTO LXI
13 Malter	15 null	nw13143 HYUNDAI VERNA
Transmits statemed 30 informed statement over informed statement over including statement and statement over informed statemen		

```
Launcher
                        X Etl_datalake.ipynb
| + % □ □ ▶ ■ C >> Code
                                                                                            Python 3 (ipykernel)
   [22]: from pyspark import SparkConf
          from pyspark.sql import SparkSession
   [23]: aws_access_key_id = 'ASIA4DKGGEBSDQZVECDU'
          aws_secret_access_key = 'H+bGZh1KAdo/AnH8V6VJHZdTh7QwSWDFUSoiKd9b'
aws_session_token = 'FwoGZXIvYXdzEEYaDLEnb8rmI/WTOC2/GCLQAeSfsbiQySq+BF93PQrmJb91syq4k12FXhn9p40U
   [24]: conf = SparkConf()
          conf.set("spark.jars.packages", "org.apache.hadoop:hadoop-aws:3.2.2")
          conf.set("spark.hadoop.fs.s3a.aws.credentials.provider", "org.apache.hadoop.fs.s3a.TemporaryAWSCr
          conf.set("spark.hadoop.fs.s3a.access.key", aws_access_key_id)
conf.set("spark.hadoop.fs.s3a.secret.key", aws_secret_access_key)
          conf.set("spark.hadoop.fs.s3a.session.token", aws_session_token)
          4 1
   [24]: <pyspark.conf.SparkConf at 0x7f01b9e35780>
   [25]: spark = SparkSession.builder.config(conf=conf).getOrCreate()
   [26]: bucket = "s3a://sarwandataset/cleaned/"
          cleaned_zone = f"{bucket}/"
   [27]: data = spark.read.option("header", "true").option("multiline", "true").csv(cleaned_zone)
          data.createOrReplaceTempView("staging_data")
   [28]: data.printSchema()
           |-- Cus_ID: string (nullable = true)
           |-- FullName: string (nullable = true)
           |-- EmailAddress: string (nullable = true)
           |-- Requested_Amount: string (nullable = true)
           |-- Emi_Amount: string (nullable = true)
           |-- Age: string (nullable = true)
           |-- Postal_Code: string (nullable = true)
           |-- Applicant_State_Desc: string (nullable = true)
           |-- Applicant_City_Desc: string (nullable = true)
            |-- Resid_Owned_By_Desc: string (nullable = true)
           -- No_Of_Years_At_Residence: string (nullable = true)
           |-- Employment_Type_Desc11: string (nullable = true)
           |-- Duration_Of_Current_Emp: string (nullable = true)
           |-- Years_In_Current_Business: string (nullable = true)
           |-- Total_Work_Experience: string (nullable = true)
           |-- Car_Register: string (nullable = true)
           |-- Product_Name1: string (nullable = true)
           |-- Product_Line: string (nullable = true)
            -- Loan_Term: string (nullable = true)
           |-- Variant_Code: string (nullable = true)
           |-- Manufacturer_Desc: string (nullable = true)
            |-- Employment_Type_Desc21: string (nullable = true)
           |-- application_creation_date: string (nullable = true)
           |-- Ex_Showroom_Price: string (nullable = true)
            -- Segment: string (nullable = true)
           |-- Segment_Desc: string (nullable = true)
            |-- Cost_Of_Vehicle: string (nullable = true)
           |-- Average_Bank_Balance: string (nullable = true)
```

Dasdasdsad



```
root

root

cus_ID: string (nullable = true)

rullMame: string (nullable = true)

regulated_Asount: string (nu
```

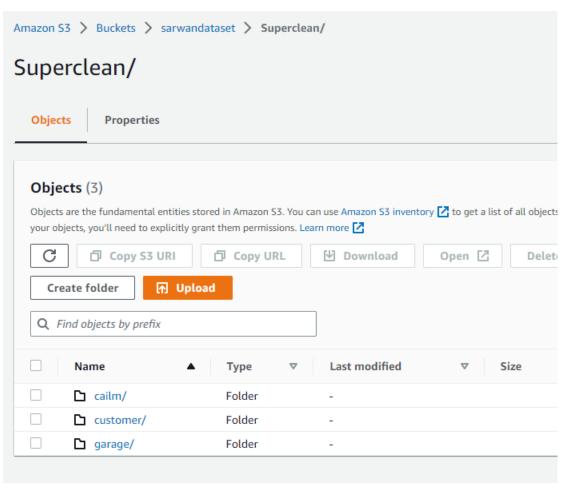
	+					+	+				+
	Zames !	Zanast@mat2_sanl	2000001		10.51 10.0001	Thombs: - 1		Proportion I	cere!	m.22.1	CALABTER
	James	James1@gmail.com	200000	5750		Thonburi	40117	Bnagkok	SELF	null SALARTED	SALARIED 4/17/2015
5	421	null	5 6		EVROLET SPARK CHEVR		48 L F 1	1.0 BS-IV OBDII CHEV	KOLET INDIA LTD	SALARIED	4/17/2015
		Compact 237342 Charles1@gmail.com	1900000	null null 4:	594 24.7 1.25 10240	101401 bangkapi		Bnagkok	SELFI	null	RETIRED
null	chartes	Charlesiggmail.com null	1900000 null	กน11 4: คมง 8688		YOTA - FORTUNER	48	FORTUNER 4 WD	SELF TOYOT		4/23/2014
	SUV	SUV 2400000	HULL	null	754 24.7	100602	40	PORTUNER + WD	10101	NEITKED]	4/23/2014
		George1@gmail.com	1600000	null		bangkapi		Bnagkok	SELFI	nul1	RETIRED
null	acai ge	null	null	กม6532		YOTA - FORTUNER	481		TOYOT		4/25/2014
	SUV	SUV 22000001	1102.2	null	779 16.5	100602	40	TONTONER STOLE	10101	KETTRED	4/23/2014
		Frank1@gmail.com	300000	98581	35 10240	bangkapi		Bnagkok	SELFI	7	SALARIED
16		null	16	คม4265	NISSAN MICRAI	NISSAN MICRA	48	NISSAN MICRA XL	NISSAN	SALARIED	1/2/2013
365750	A21	Compact 365750		nul1	818 18.1	100602					-, -, -323
	Joseph		2500001	8750	34 10240	bangkapi		Bnagkok	SELFI	2 SELF EMPL	LOYED NON
null		3	null	คม7067	XYLO E8	XYLO E8	36	XYLO E8		AL SELF EMPLOYED NON	
5487501	MUV	MUV 548750		nul1	853 17.69	100602	- 10				-,, 2025
		Thomas1@gmail.com	150000	5127	33 10240	bangkapi		Bnagkok	SELF	25 SELF EMPI	LOYED NON
null		6	null	กม10060	FORD FIGO	FORD FIGO	361	FORD FIGO 1.4 LXI		A SELF EMPLOYED NON	1/17/2013
329250	A2	Compact 329250		nul1	777 14.66	100602					-, -, /
7	Henry	Henry1@gmail.com	430000	14592	45 10240	bangkapi		Bnagkok	SELF	40 SELF EMPI	LOYED PRO
nul1		15	nul1	คม12815	MARUTI SWIFT	MARUTI SWIFT	361	SWIFT DZIRE LDI		SELF EMPLOYED PRO	1/21/2013
482750	A3	Mid-Size 482750		nul1	817 13.86	100602					
8	Robert	Robert1@gmail.com	405000	13842	46 10240	bangkapi		Bnagkok	SELF	9 SELF EMPI	LOYED NON
nul1	-1	10	nu11	กม11055	MARUTI RITZ	MARUTI RITZ	36	RITZLDI		SELF EMPLOYED NON	
461500	A2	Compact 461500		nul1	789 14.37	100602					
9	Edward	Edward1@gmail.com	276000	9567	34 10240	bangkapi		Bnagkok	SELF	10 SELF EMPI	LOYED NON
nul1		5	nul1	กม10777	HYUNDAI VERNA	HYUNDAI VERNA	36	VERNA CRDI SX	HYUNDA	[SELF EMPLOYED NON	1/31/2013
437250	A3	Mid-Size 437250		nul1	828 14.54	100602					
10	Harry	Harry1@gmail.com	234000	7997	34 10240	bangkapi		Bnagkok	SELF	25 SELF EMPI	LOYED NON
nul1		5	nul1	คม12420	ALTO LXI	MARUTI ALTO LXI	36	ALTO LXI	MARUT	[SELF EMPLOYED NON	2/12/2013
167737	A2	Compact 167737		nul1	837 14.14	100602					
11	Walter	Walter1@gmail.com	610000	20847	36 10240	bangkapi		Bnagkok	PARENTS	10 SELF EMPI	LOYED NON
nul1		15	null	คม13143	HYUNDAI VERNA	HYUNDAI VERNA	36	HYUNDAI VERNA XXI	HYUNDA	I SELF EMPLOYED NON	2/13/2013
868500	A3	Mid-Size 1737000		nul1	738 14.03	100602					
	Arthur	Arthur1@gmail.com	200000	nul1	37 10240	bangkapi		Bnagkok		null SELF EMPI	
nu11		nul1	nu11	คม15814	HYUNDAI I 10	HYUNDAI I 10	36	I-10	HYUNDA	I SELF EMPLOYED NON	2/16/2013
null n	u11	null 243250		null	-1 16.2	100602					
14	Fred	Fred1@gmail.com	300000	11160	58 10240	bangkapi		Bnagkok	SELF	10 SELF EMPI	
nul1		30	nul1	กม5913	MARUTI SWIFT	MARUTI SWIFT	36	SWIFT VDI	MARUT	I SELF EMPLOYED NON	2/14/2013
345750	A2	Compact 345750		nul1	788 14.1	100602					
	Albert	Albert1@gmail.com	292000	9980	47 10240	bangkapi		Bnagkok	SELF	15 SELF EMPI	
nul1		15	nul1	คม6814	SKODA FABIA	SKODA FABIA	36	FABIA AMB 1.2	SKOD	A SELF EMPLOYED NON	2/23/2013
356000	A3	Mid-Size 356000		nul1	518 17.3	100602					
	Samuel	Samuel1@gmail.com	437500	14952	38 10240	bangkapi		Bnagkok	SELF		LOYED NON
nul1		10	nu11		HYUNDAI - ACCENT	HYUNDAI - ACCENT	36	ACCENT EXECUTIVE	HYUNDA	I SELF EMPLOYED NON	2/26/2013
312500	A3	Mid-Size 312500		nul1	793 22.5	100602					
17	David	David1@gmail.com	1260000	43065	63 10240	bangkapi		Bnagkok		20 SELF EMPI	
null		5	nul1	คม6747	CHEVROLET CRUZE	CHEVROLET CRUZE	36	CHEVROLET CRUZE	CHEVROLET INDIA LT	D SELF EMPLOYED NON	2/27/2013
1100000	A4	Executive 1100000		nul1	-1 14.61	100602					
18	Louis	Louis1@gmail.com	1240000	42380	38 10240	bangkapi		Bnagkok	SELF	20 SELF EMPI	
nul1		12	null	กต.8688	SKODA YETI	SKODA YETI	36	SKODA YETI	SKO	A SELF EMPLOYED NON	2/27/2013
1	1150000		1150000		null		00602	200			
19	Joe	Joe1@gmail.com	220000	10780		bangkapi		Bnagkok	SELF		LOYED NON
nul1	470000	16	null	nai6532		MARUTI ALTO	36	ALTO LX	MARU	I SELF EMPLOYED NON	3/5/2013
1	170000		170000		null		80602			-1	
	Charlie	Charlie1@gmail.com	1198000	40945		bangkapi		Bnagkok	SELF		LOYED NON
nul1	4077000	20	null	คต.4265		SKODA SUPERB	36	SKODA SUPERB 1.8	SKO	A SELF EMPLOYED NON	3/5/2013
1 2414	1877250		1877256		null		00602	Providental Control	DELIZED I	21 55 5 500	auto non I
		larence1@gmail.com	255000	12304	38 10240	bangkapi		Bnagkok	RENTED		LOYED NON
nul1	357886		nu11 357006	คดเ7067	FIAT PUNTO null	FIAT PUNTO 763 14.46 18	24 99692	PUNTO DYNAMIC DI	FIAI INDIA L	D SELF EMPLOYED NON	3/7/2013
1	357000	A2 Compact	35/006	9	until	/03[14.46] 18	90002				
+											

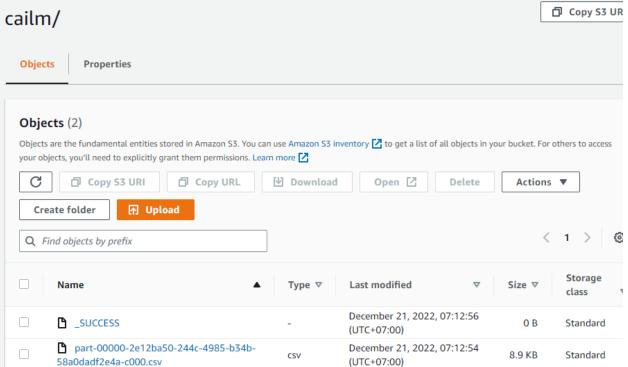
14]: data.select("Age", "Product_Name1").show()

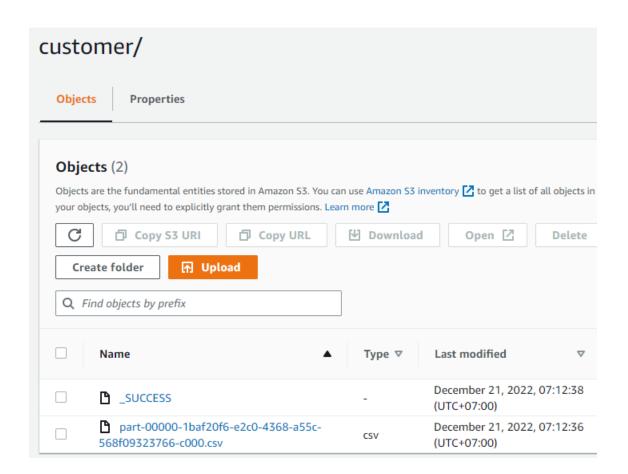
Ι Δσεί	Product_Name1						
++							
40.5	CHEVROLET SPARK						
41.25	FORTUNER 4 WD						
39.5	null						
35	NISSAN MICRA						
34	XYLO E8						
33	FORD FIGO						
45	MARUTI SWIFT						
46	MARUTI RITZ						
34	HYUNDAI VERNA						
34	ALTO LXI						
36	HYUNDAI VERNA						
37	HYUNDAI I 10						
58	MARUTI SWIFT						
47	SKODA FABIA						
38	HYUNDAI - ACCENT						
63	CHEVROLET CRUZE						
38	SKODA YETI						
57	MARUTI ALTO						
40	SKODA SUPERB						
38	FIAT PUNTO						
++							
only sh	owing top 20 rows						

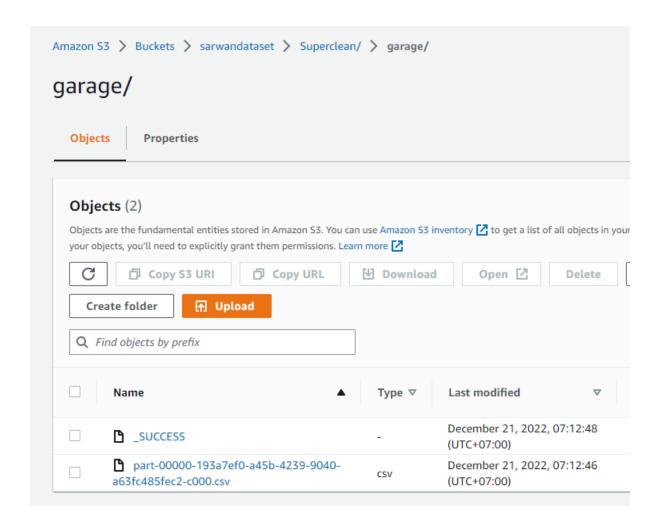
+----+

```
table_customer = spark.sql("""
                                                         □ ↑ ↓ 古 무 🗎
  select
     Cus_ID
     ,FullName
     ,EmailAddress
     ,Requested_Amount
     ,Age
     ,Applicant_State_Desc As State
     ,Applicant_City_Desc As City
     ,Employment_Type_Desc11
     ,Total_Work_Experience
     ,Product_Name1
     ,Loan_Term
     ,Cost_Of_Vehicle
     ,Insurance_Code
   from staging_customer where Product_Name1 is not null
table_customer.createOrReplaceTempView("customer")
table_garage = spark.sql("""
   GarageID
   ,name
,PhoneNumber
    ,Address
 from stagging_garage
table_garage.createOrReplaceTempView("garage")
table_cailm = spark.sql("""
   select
   from stagging_cailm
table_cailm.createOrReplaceTempView("cailm")
table_customer.write.mode("overwrite").option("header",True).csv(exportData+"/customer")
table_garage.write.mode("overwrite").option("header",True).csv(exportData+"/garage")
table_cailm.write.mode("overwrite").option("header",True).csv(exportData+"/cailm")
```









Step3

Data Warehouse zone เป็นส่วนที่ใช้จัดเก็บไฟล์ผ่านการ transform หรือ ETL ให้อยู่ใน postgrest เนื่องจาก aws redshift ติด access denied