

Project Instruction

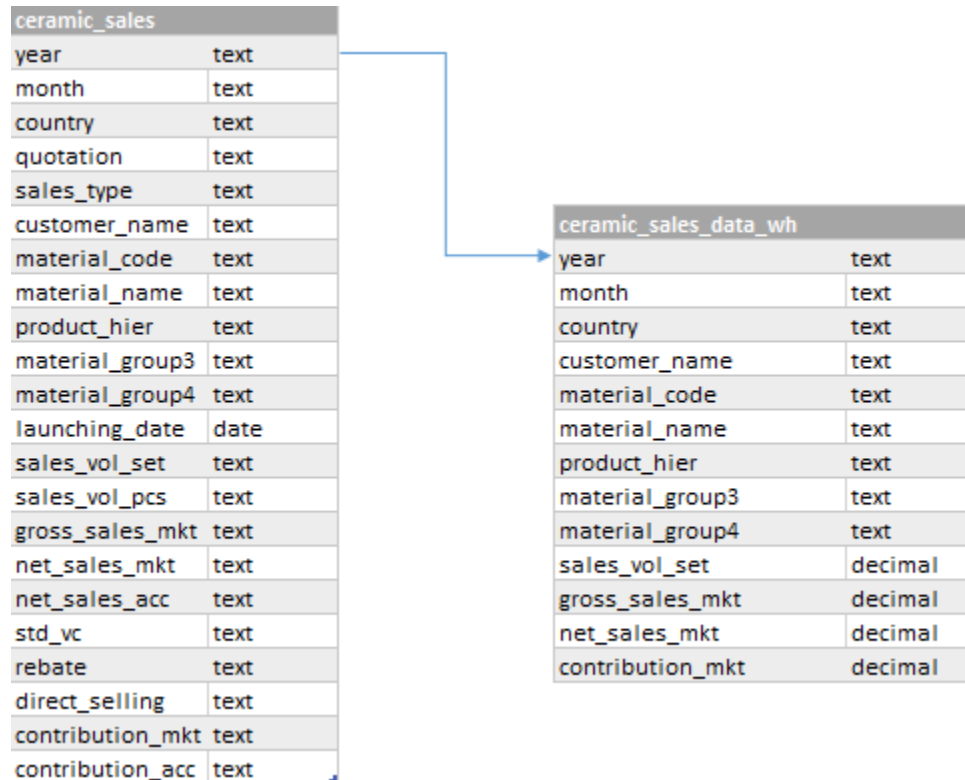
บริษัท ไข่ทองคำ จำกัด เป็นธุรกิจผลิตและจัดจำหน่ายวัสดุก่อสร้าง อุปกรณ์ ใน
ห้องน้ำซึ่งมียอดขายอันดับหนึ่งในประเทศไทย ต้องการผลักดันยอดขายในต่างประเทศ
มากขึ้น แต่เนื่องจากว่าแผนกขายต่างประเทศมีลูกค้าอยู่ในหลายประเทศ และแต่ละ
ประเทศก็มีการขายสินค้าจำนวนมากที่แตกต่างกัน ทำให้จัดการแผนการตลาดได้
ค่อนข้างยาก ทางแผนกต้องการแก้ปัญหาโดยได้วิเคราะห์ข้อมูลดังนี้

1. ดูตลาดหลักของการขายสินค้า
2. ดูความสัมพันธ์ระหว่างยอดขายและกำไร
3. ดูสินค้าที่สามารถทำตลาดร่วมกันได้ในหลายประเทศ

แหล่งที่มาของข้อมูล – ยอดขายต่างประเทศตั้งแต่ปี 2018-2022 ของบริษัท ไขทองคำ จำกัด (ชื่อไฟล์ Ceramic_sales.csv)

[illegible]

Data Model



มีทั้งหมด 2ตาราง โดยการสร้างตารางใน **data warehouse** จะเป็นการดึงมาเฉพาะข้อมูลที่เกี่ยวข้องและจำเป็นในการวิเคราะห์ข้อมูล

Project workflow

Cloud access AWS

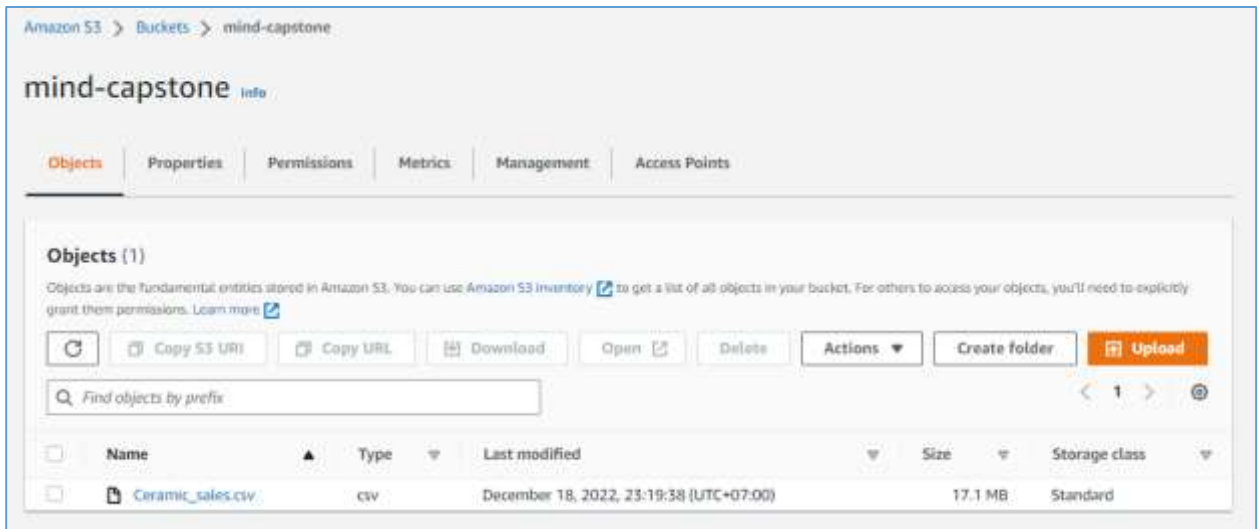
1. Get Access Key จาก AWS

- `$ cat ~/.aws/credentials`

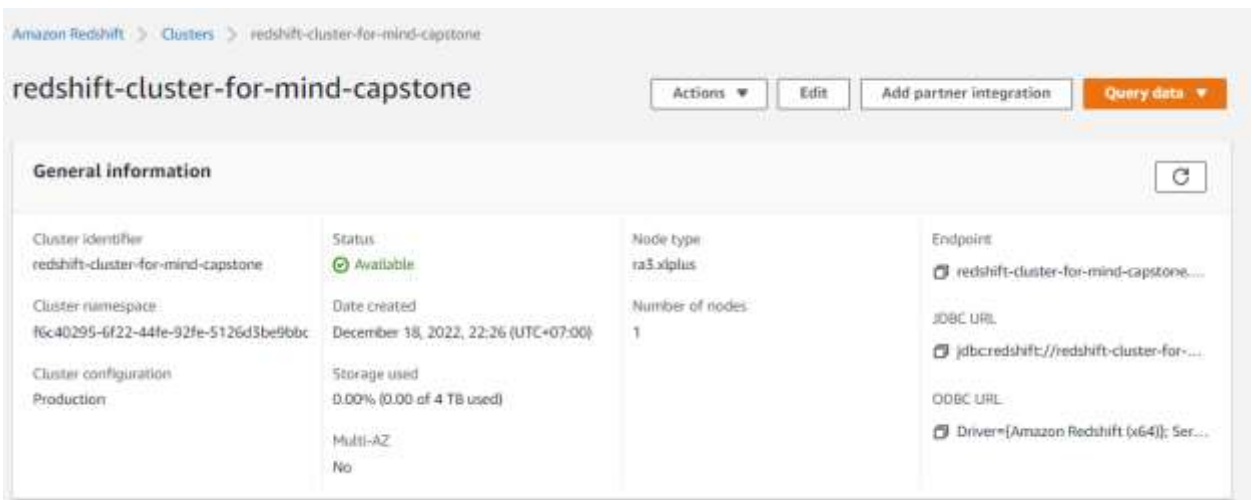
```
ddd_v1_w_GPj_1383621@runweb69692:~$ cat ~/.aws/credentials
[default]
aws_access_key_id = ASIAQ6I4WHZWH333507Z
aws_secret_access_key = y3V6X3KWH1L1Km1TDpWk70pJ2N00B7AHJHGHMFX
aws_session_token = FwoGZKIvYXdeEFu0DPKegX61gRV5nRPKCLIAWH0H3Aue/jkh0QXfmd4p0U5mneKco3hh0MF2iF4McH0u2G93iR
6Iuvzn/T8u1N0qSOATwTZnbpcPukGmkvqZHuoA7U51c7c1m216YpSGhtn81ZgXw5IWsYB/M/No/b5/WfJ0n3cxwccThccPfwuGwoIa641ox
30ND7pSzIHb0hh1AFlyfag6a/iFYk0pTHX0s8j6cCvZPTGYAGVqH1-CapLawnSaQ28eMcHjq-HubwymiUtsNv4uhQFcF17FHh1nk1u2vPVK1L
KLqT130GM13pQ2xptcG-0mm6ESM40QnHp-e31Vrr3vE5GH5qAR2d4VLOH40n8T30NoNyGhg
ddd_v1_w_GPj_1383621@runweb69692:~$
```

- จะได้ 3 keys ดังนี้เพื่อนำไปใส่ใน code บน Github เพื่อเชื่อมต่อ AWS
 - ✓ aws_access_key_id
 - ✓ aws_secret_access_key
 - ✓ aws_session_token

2. นำข้อมูลขึ้นไปไว้ใน S3 โดยเปิดให้เป็น public



3. สร้าง Redshift โดยให้ Node type เป็น ra3.xplus และมีจำนวน 1node ตั้ง username และ password ให้เรียบร้อย และเปิดให้เป็น public



Github

1. Change directory

- `$ cd 08-capstone-project`

2. สร้าง virtual environment ชื่อ "ENV" (ในครั้งแรก)

- `$ python -m venv ENV`

3. Activate เข้าสู่ environment

- `$ source ENV/bin/activate`

4. ติดตั้ง libraries ที่จำเป็น (ในครั้งแรก)

- `$ pip install -r requirements.txt`

5. เตรียม code ในไฟล์ etl.py เพื่อนำเข้าข้อมูล load และ transform

6. เตรียมติดตั้งและเปิด Docker

- `$ mkdir -p ./dags ./logs ./plugins`
- `$ echo -e "AIRFLOW_UID=$(id -u)" > .env`
- `$ docker-compose up`

7. เปิด Airflow

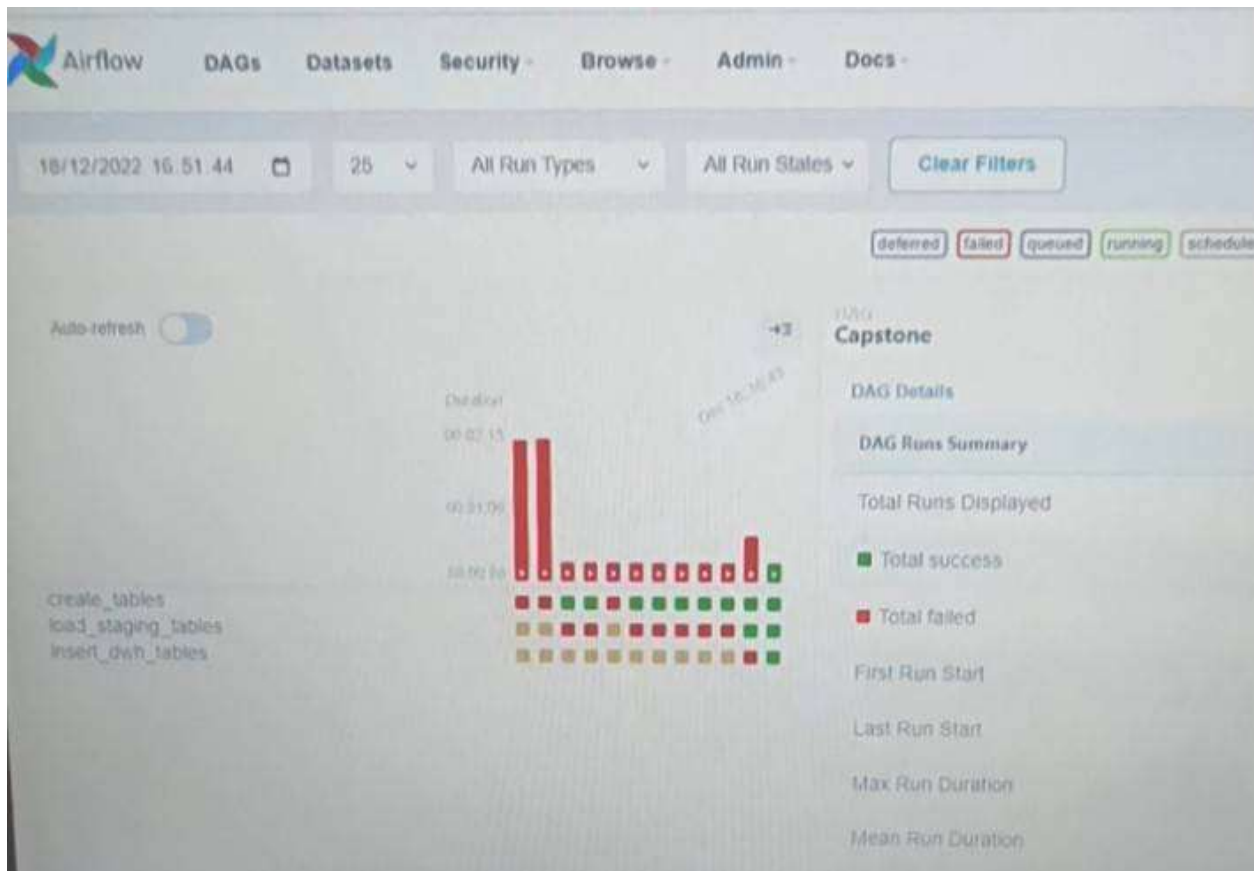
- Port 8080

8. เข้า DAGs ที่ชื่อว่า Capstone

9. สร้าง connection

- Connection id: my-redshift
- Connection Type: Redshift
- Host: มาจาก Endpoint บน Redshift
- Schema: dev
- Login: awsuser
- Password คือ Password ที่ตั้งตอนสร้าง Redshift
- Port: 5439

10. RUN DAGs ชื่อ “Capstone”



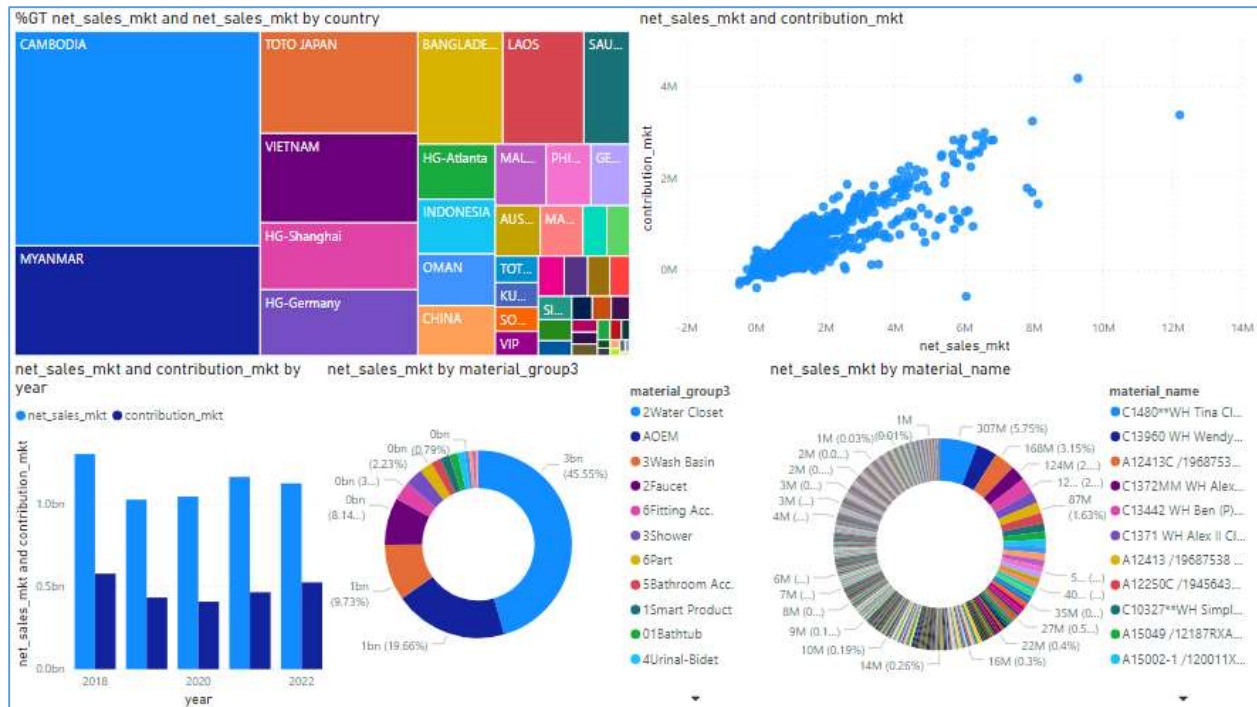
Cloud access AWS

1. หลังทำการ ETL ข้อมูลเรียบร้อยแล้ว ให้เข้าไปที่ Redshift เพื่อเชื่อมต่อกับ Database
2. ทำการ Export CSV Files ออกมาเพื่อนำไปใช้งานต่อไป

Github

1. Shutdown Docker ปิดการทำงานของ Docker
 - \$ docker-compose down
2. Deactivate the virtual environment
 - \$ deactivate

Data Visualization by PowerBI



1. ตลาดหลักของขายต่างประเทศคือ กัมพูชา รองลงมาคือ พม่า
2. ความสัมพันธ์ระหว่างยอดขายและกำไรค่อนข้างเป็นไปในทิศทางเดียวกัน แต่มีจุดที่น่าสังเกตคือจากปี 2019 ไป 2020 ยอดขายเพิ่มขึ้นแต่กำไรลดลง ส่วนในปี 2021 ไป 2022 ยอดขายลดลง แต่กำไรเพิ่มขึ้น อาจเกิดจากต้นทุนของสินค้าที่มีความผันแปร ซึ่งในส่วนนี้ก็ต้องรายงานไปยังผู้จัดการเพื่อให้หารือกับแผนกที่เกี่ยวข้องต่อไป
3. สินค้าที่ขายมีความหลากหลายมากจนเกินไปทำให้ไม่สามารถหาโมเดลสินค้าที่ร่วมกันได้ทุกประเทศ แต่อาจทำเป็นโซนประเทศใกล้เคียงกันมารวมกันได้ เช่น กัมพูชา พม่า และลาว