Kelompok 13

Final Project Presentation

Modeling And Optimization Techniques In Data Warehousing Bootcamp

Profile Team

Adnanfadhil Yaser

Universitas Pembangunan Nasional Veteran Yogyakarta

Fajar Mulia Ananda

Universitas Gunadarma

Sela Aziza

Universitas Muhammadiyah Kotabumi

Putri Sitti Naima

Institut Teknologi Sepuluh Nopember

Background

Data Engineer!

US online retail company menjual produk pelanggan umum langsung ke pelanggan dari berbagai pemasok di seluruh dunia. Tantangannya adalah membangun infrastruktur data menggunakan data yang dihasilkan yang dibuat untuk mencerminkan data dunia nyata dari perusahaan. Inilah tugasnya :

- **ETL/ELT Job Creation menggunakan Airflow**
- **Data Modeling di Postgres**
- **Dashboard Creation dengan Data Visualisasi**

Stack

1

Data

Mendapatkan 5 format data yang berbeda

2

Airflow

Airflow DAG and mempersiapkan ETL Data

3

Data Modelling

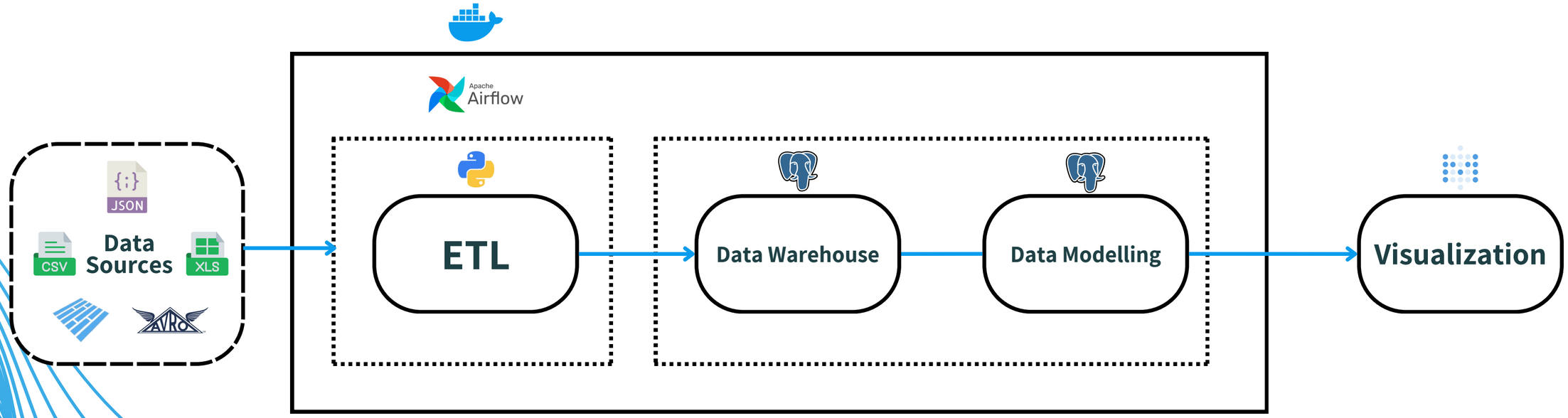
Struktur dari data modelling

4

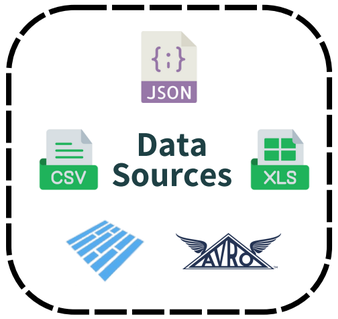
Visualisasi

Visualisasi data yang telah diproses

Data Pipeline



Data



requirements.txt

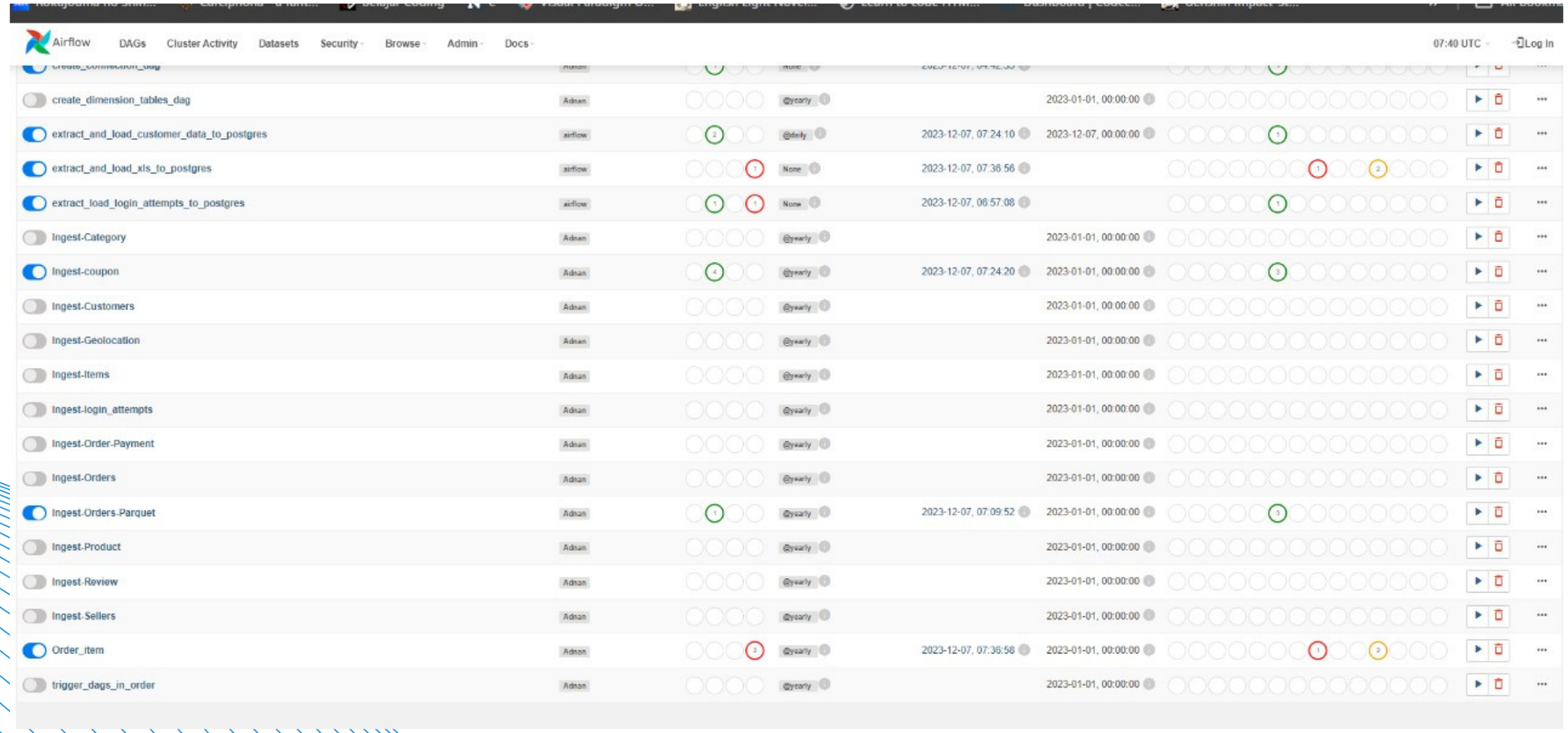
Dibimbing_FinalTest-main > requirements.txt

```
pandas==2.1.0
openpyxl
fastparquet==0.8.2
fastavro==1.4.6
avro-python3==1.10.2
pyarrow==5.0.0
psycopg2-binary==2.9.1
xlrd==1.2.0
python-snappy==0.6.0
```

```
from datetime import datetime, timedelta
from airflow import DAG
from airflow.providers.postgres.hooks.postgres import PostgresHook
from airflow.operators.python import PythonOperator
import pandas as pd
from sqlalchemy import create_engine
import pyarrow.parquet as pq
import fastavro
import os
import json
```

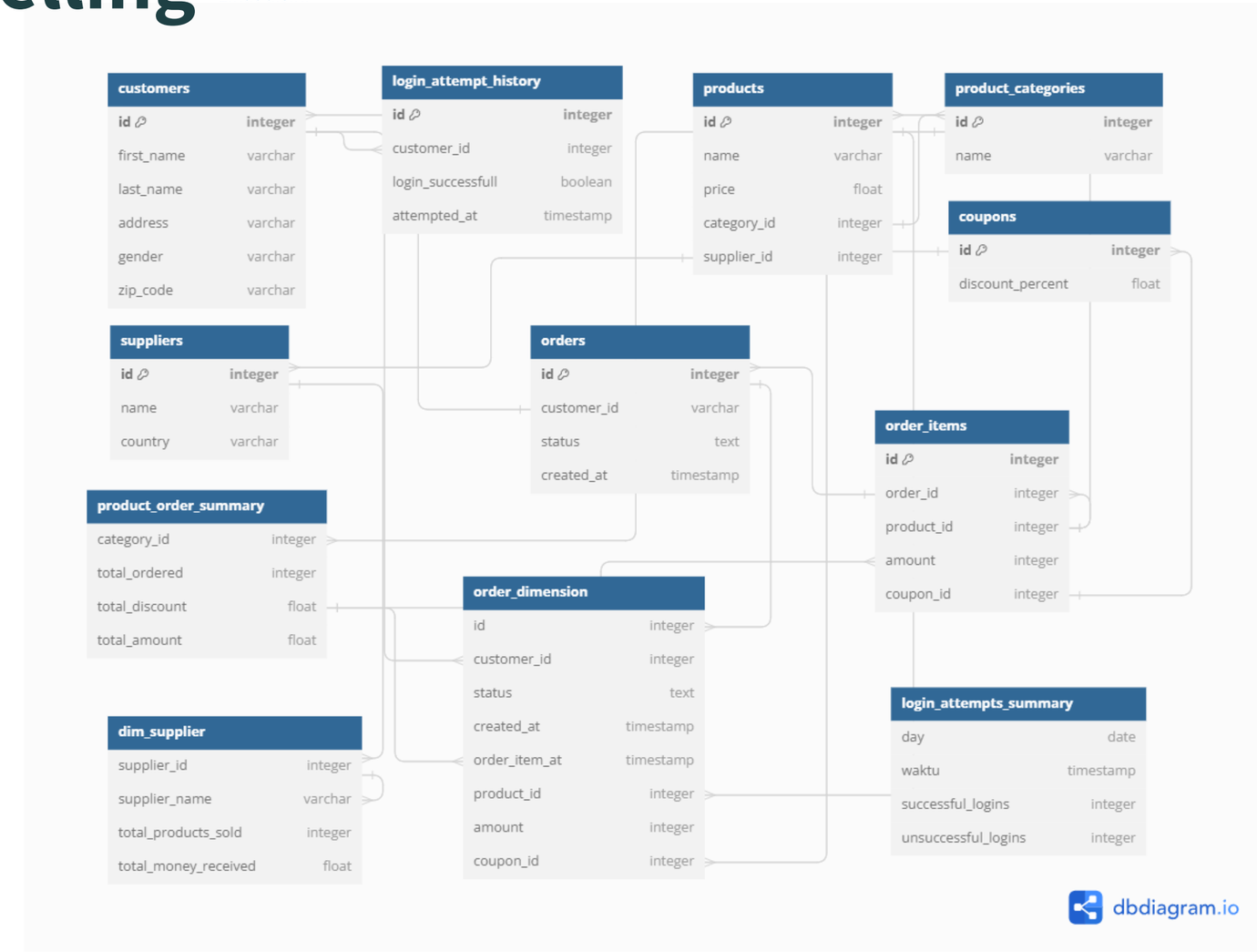
Yang dibutuhkan untuk mengolah lima data dengan format yang berbeda

Airflow


Airflow
DAGs
Cluster Activity
Datasets
Security
Browse
Admin
Docs
07:40 UTC
Log In
create_dimension_tables_dag
extract_and_load_customer_data_to_postgres
extract_and_load_xls_to_postgres
extract_load_login_attempts_to_postgres
Ingest-Category
Ingest-coupon
Ingest-Customers
Ingest-Geolocation
Ingest-Items
Ingest-login_attempts
Ingest-Order-Payment
Ingest-Orders
Ingest-Orders-Parquet
Ingest-Product
Ingest-Review
Ingest-Sellers
Order_item
trigger_dags_in_order

Data Modelling

Lihat disini



ibimbing

Data Modelling

Lihat disini

product_order_summary	
category_id	integer
total_ordered	integer
total_discount	float
total_amount	float

Untuk melihat seberapa baik produk terjual, kategori apa yang paling diminati

order_dimension	
id	integer
customer_id	integer
status	text
created_at	timestamp
order_item_at	timestamp
product_id	integer
amount	integer
coupon_id	integer

Untuk melacak detail pesanan, pelanggan.

dim_supplier	
supplier_id	integer
supplier_name	varchar
total_products_sold	integer
total_money_received	float

Untuk menilai pemasok berkinerja dan berapa banyak uang yang diperoleh dari penjualan produk pemasok.

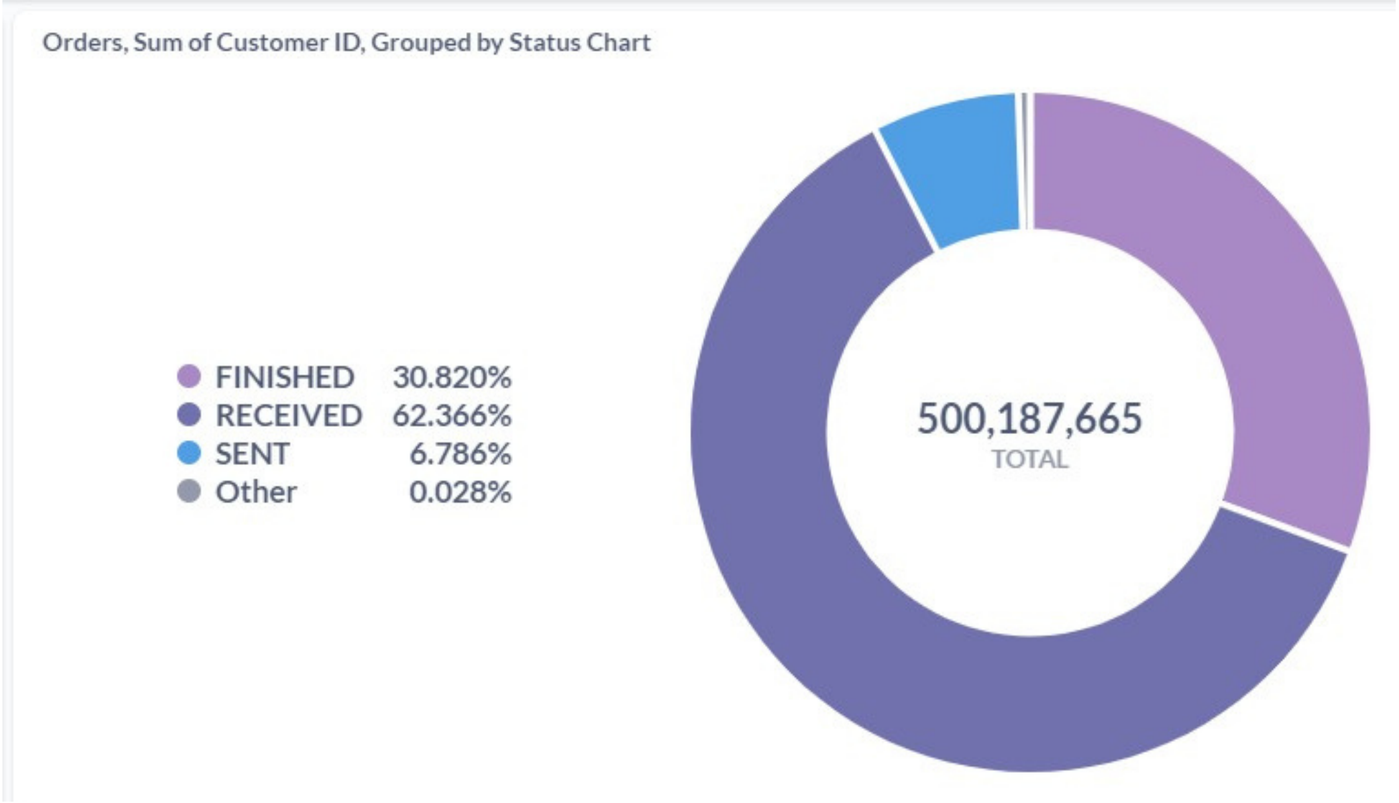
login_attempts_summary	
day	date
waktu	timestamp
successful_logins	integer
unsuccessful_logins	integer

Untuk melihat pola aktivitas login

Visualisasi

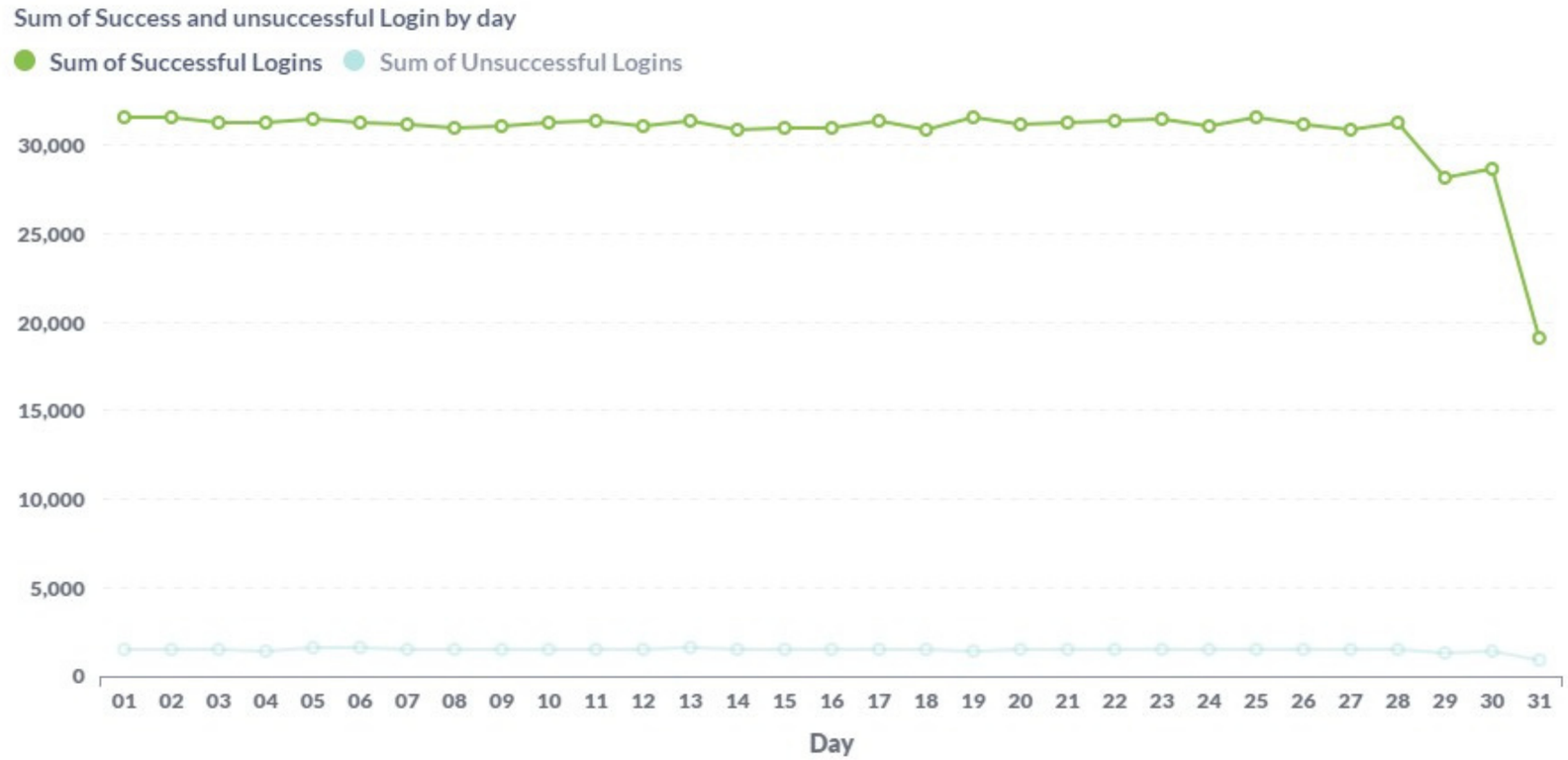
Hasil dari visualisasi terlihat bahwa kategori ‘food’ memiliki jumlah total order yang terbanyak mencapai $18.8M, dan diikuti kategori ‘fashion’ mencapai $ 12.5M. Kedua kategori ini yang memiliki pengaruh pendapatan pada perusahaan ini.

Visualisasi



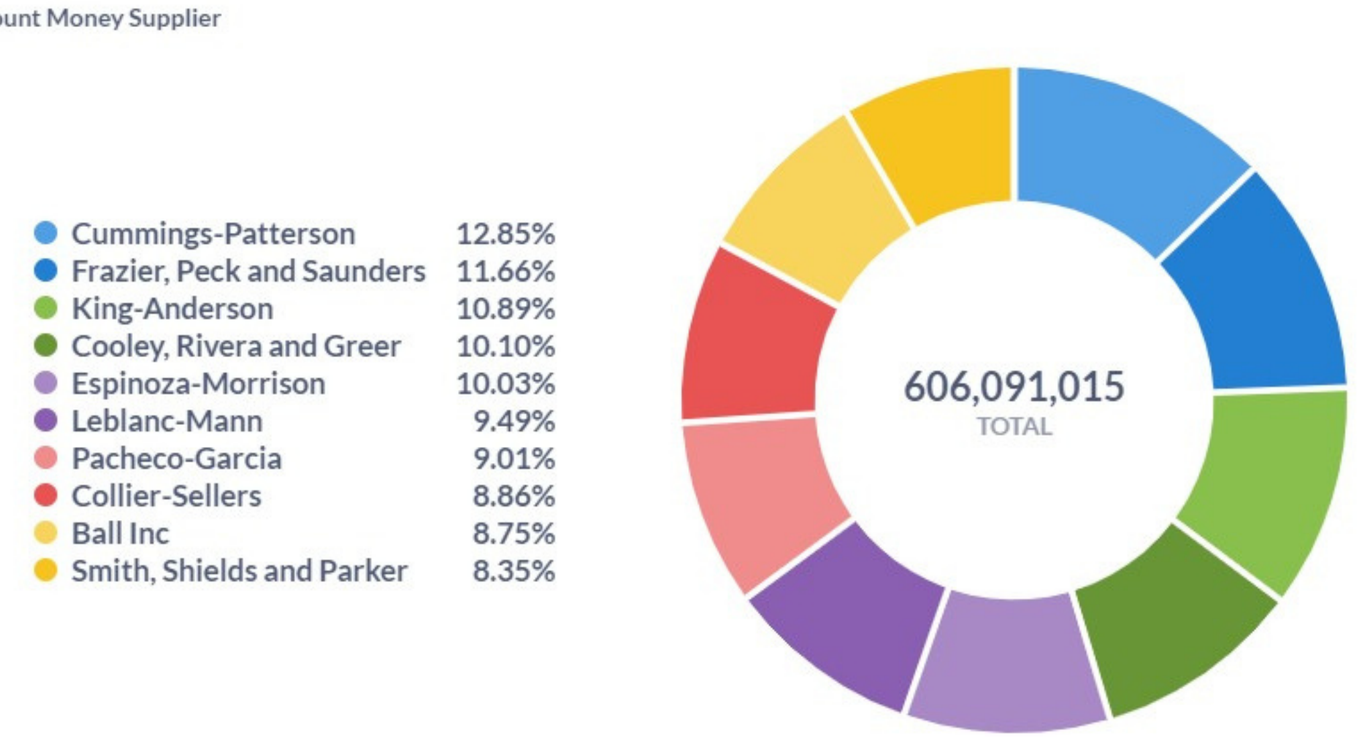
Hasil dari visualisasi terlihat bahwa detail pesanan pada bulan tersebut mayoritas sudah diterima dan selesai oleh pelanggan yaitu sebesar 93%.

Visualisasi



Hasil dari visualisasi terlihat bahwa siklus dari berhasil login dan tidak berhasil mengalami naik turun berdasarkan perilaku pengguna dalam menggunakannya.

Visualisasi



Berdasarkan grafik 10 pemasok dana teratas, dapat disimpulkan bahwa Cummings-Patterson adalah pemasok dana terbesar, dengan pangsa aset sebesar 12,85%, diikuti oleh Frazier, Peck dan Saunders dengan pangsa aset sebesar 11,66%, dan King-Anderson dengan pangsa aset sebesar 10,89%.

Thank You

Kelompok 13