Data Warehouse Project

Project Title: "Analyzing Customer Reviews of Bank Agencies in Morocco using a Modern Data Stack"

Objective:

The goal of this project is to **collect, process, and analyze Google Maps reviews** for bank agencies in Morocco to extract valuable insights using topics analysis, sentiment detection, and other key insights.

Students will build a **fully operational data pipeline** using modern tools, ensuring efficient **data extraction, transformation, storage, and visualization.**

1. Project Scope

🔹 Use Case Description

Banks receive thousands of customer reviews on Google Maps. These reviews contain **valuable insights** about customer satisfaction, service quality, and common issues. However, this data is **unstructured and scattered** across different locations. The aim of this project is to **centralize, clean, and analyze** this data to provide meaningful insights.

Expected Insights:

1. **Sentiment Analysis:** Understanding customer satisfaction trends.
2. **Topic Modeling:** Identifying common issues and praise points.
3. **Branch Performance:** Ranking bank agencies based on customer sentiment.
4. **Customer Experience Metrics:** Extracting recurring complaints or satisfaction drivers.

2 Tech Stack

Stage	Technology
Data Collection	Python, Google Maps API, BeautifulSoup/Scrapy
Scheduling	Apache Airflow
Data Storage	PostgreSQL (Data Warehouse)

Stage	Technology
Transformation	DBT (Data Build Tool)
Analysis & BI	Looker Studio (Google Data Studio)
Version Control	GitHub

3 Project Roadmap

Students will complete the project in **6 main phases**:

✅ Phase 1: Data Collection (Scraping Google Maps Reviews)

1. **Use Google Maps API or Web Scraping**
    - Extract customer reviews for all major bank agencies in Morocco.
    - Data includes: Bank name, branch name, location, review text, rating, review date.
    - Store raw data in JSON/CSV format.
2. **Automate with Apache Airflow**
    - Schedule a DAG to collect data daily/weekly.
    - Store extracted data in a PostgreSQL staging table.

✅ Phase 2: Data Cleaning & Transformation

1. **Clean the Data (DBT & SQL)**
    - Remove duplicate reviews.
    - Normalize text (lowercase, remove punctuation, stop words).
    - Handle missing values.
2. **Enrich the Data**
    - Extract language from review text.
    - Classify reviews into "Positive," "Negative," "Neutral" using sentiment analysis.
    - Extract common topics using NLP (Latent Dirichlet Allocation - LDA).

✅ Phase 3: Data Modeling (Star Schema in PostgreSQL)

1. **Design a Data Mart using Star Schema**
    - **Fact Table:** fact_reviews
    - **Dimension Tables:**
        - dim_bank
        - dim_branch
        - dim_location
        - dim_sentiment
2. **Load Data into PostgreSQL**
    - Implement transformation models in **DBT**.
    - Automate data ingestion with Airflow.

✅ Phase 4: Data Analytics & Reporting

1. **Build Dashboards in Looker Studio**
    - Sentiment trend per bank & branch.
    - Top positive & negative topics.
    - Performance ranking of branches.
    - Customer experience insights.

✅ Phase 5: Deploy & Automate

1. **Automate the Entire Pipeline**
    - Use **Airflow** to schedule daily/weekly updates.
    - Implement **alerts for data failures**.

---

4 Deliverables

Students will submit:

1. **Python script** for data collection.
2. **Airflow DAGs** for automation.

3. **DBT models** for data transformation.
4. **PostgreSQL data warehouse schema**.
5. **Looker Studio Dashboard**.
6. **Project Documentation (README file).**

5 Grading Criteria

Category	Points
Data Collection (Google Maps API or Scraping)	20%
Automation (Airflow DAGs)	20%
Data Transformation (DBT & SQL)	20%
Data Modeling (PostgreSQL - Star Schema)	20%
BI Dashboard (Looker Studio)	20%

🔹 **Summary of Project Timeline**

Week	Focus Area	Key Deliverables
Week 1	Data Collection (Scraping)	Google Maps API/Scraping, Airflow DAG, Raw Data in PostgreSQL
Week 2	Data Cleaning & Transformation	DBT models, Sentiment Analysis, Airflow Update
Week 3	Data Modeling (Star Schema)	Fact & Dimension Tables, SQL Scripts, Data Loading
Week 4	Analytics & BI Dashboards	Looker Studio Dashboards, Topic Modeling
Week 5	Final Report & Presentation	Documentation, GitHub Repo, Presentation