Capstone Project

Problem Statement:

Enabling More Targeted Promotions and Lower Customer Acquisition Costs

One of Incedo Inc banking and financial services (BFSI) customer wanted to improve their marketing campaigns in order to boost conversion rates while lowering customer acquisition costs. To better target clients and promote the most appropriate products and services, the bank sought to identify channels, offers, and approaches.

Data Description:

The bank decided to collate, aggregate, and analyze a large amount of structured and unstructured data to help identify key indications of interest, including

1. Customer purchase and transaction history
2. Customer profile data
3. Customer behaviour on social media (Twitter, Facebook)

Data Schema:

1. **Customer purchase and transaction history**

Attribute	Type	Format	Sample
customerNumber (CIF)	Numeric		12321312
cardLast4Digits	String		1234
transactionDate	Date	mm-dd-yyyy	11-20-2013
transactionTime	Time	hh:mm:ss	15:35:42
transactionAmount	Numeric		$234.23
transactionType	String		Swipe

merchantName	String		Walmart
merchantCity	String		Brandon
merchantState	String		FL
merchantZip	String		Single
stateCode	String		HR

2. **Customer profile data**

Attribute	**Type**	**Format**	**Sample**
customerNumber (CIF)	Numeric		12321312
dateOfBirth	Date	dd-mm-yyyy	1989-10-15
branchNumber	String		BR101
citizenshipCode	String		IND
currCountryCode	String		IND
employmentStatus	String		Salaried
incomeInThousands	Numeric		1000000
marketSegment	String		Standard
maritalStatus	String		Single
stateCode	String		HR

City	String		Gurugram
customerOpeningDate	Date	mm-dd-yyyy	12-05-2012
Gender	String		Male

3. Social Media: Twitter, Facebook

    1. Twitter Data Dictionary
    2. Twitter Data Entities
    3. Facebook APIs, SDKs & Guides - Facebook Developer Docs

TASK:

Develop a data pipeline to process a large volume of financial transaction data, customer's social media data and apply rules-based and AI-based algorithms. Innovative digital marketing campaigns services can only be possible when banks and financial services providers have the right infrastructure in place, defined by qualities that include:

1. Cloud-native and agile
2. Built for streaming and able to handle burst capacity
3. Integrated with robust security features
4. Able to provide data insights through AI
5. API-enabled

Recommended Tech Stack:

1. Cloud: AWS
2. Data Storage: AWS S3/ Redshift
3. Data Engineering ETL/ELT: Spark(PySpark) / Databricks / Glue
4. Real Time data ingestion: Amazon Managed Streaming for Apache Kafka (Amazon MSK)
5. Orchestration: Amazon Managed Workflows for Apache Airflow (MWAA)

Final Deliverables:

A word document or PPT that should contain a description of what you have tried to build.

a. What problem you are solve
b. Solution approach

c. Any supporting assumptions, functional requirements(FR) and non-functional requirements(NFR)
d. List of AWS services used with reasoning
e. Source code - Submit your Sample Code snippets used for EDA (This is applicable only for EDA. Feel free to use open source for the best approach)
f. Reason why your solution should be considered for the final round

Judging Criteria:

1. Usefulness and Completeness 25%
2. Methodological soundness or Innovative approach 25%
3. Replicability and Performance on (volume, variety, velocity, and veracity of data) 20%
4. Novelty or Originality 15%
5. Quality of presentation, and quantitative analysis when applicable being preferred 15%