Final Project Proposal

Cong Wang 001002178

1. OVERVIEW

Currently, there are lots of marketers who still insist on selling their products directly via the mail. They send catalogs with product characteristics to customers who then directly from the catalogs. Therefore, we want to learn that marketers have developed customer records to learn what makes some customers spend more than others and predict their spending.

2. GOAL

According to our dataset, we would solve these questions:

(1) The customer's distribution including their age, gender, salary, family condition
(2) Do the analysis on spending conditions based on age group, gender group etc.
(3) Describe the importance of metrics for the target column (spending)
(4) Connect the SnowFlake and build models
(5) Build the Streamlit application to demonstrate the predictions results.

3. DATASET

Dataset: Direct Marketing dataset on Kaggle
Links: https://www.kaggle.com/datasets/yoghurtpatil/direct-marketing

Here are columns:
Age: Age of customer (old/middle/young).
Gender: Gender(Cinsiyet) (Male/ Female)
OwnHome: Whether customer owns home Married(single/married)
Location: In terms of distance to the nearest brick and mortar store that sells similar products
Salary: Yearly salary of customer in dollars
Children: Number of children (0-3)
History: History of previous purchase volume (low/medium/high/NA NA means that this customer has not yet purchased)
Catalogs: Number of catalogs sent
AmountSpent: Amount Spent in dollars

4. Process Outline

The steps would be:
(1) The customer's distribution including their age, gender, salary, family condition
(2) Do the analysis on spending conditions based on age group, gender group etc.
(3) Describe the importance of metrics for the target column (spending)
(4) Design the SnowFlake configuration
(5) Building the model with RandomForest and Linear Regression from SnowFlake
(6) Building the streamlit application to demonstrate the predictions results.

5. Milestone

Timeframe	Delivery
Day 1-2	Data processing and Exploratory Data Analysis
Day 3-5	Model Building, Training, Selection
Day 6-8	Connect and Deployment Snowflake, build

	streamlit
Day 9-10	Documentation and Presentation preparation

6. Persons

Cong Wang

7. Deployment Details

Language: Python and SQL
Cloud tools and Platforms: AWS, SnowFlake
Tools for Analysis: Jupyter notebook
Visualization Tools: Streamlit

8. Reference and Resources

https://www.kaggle.com/datasets/yoghurtpatil/direct-marketing
https://www.kaggle.com/datasets/yoghurtpatil/direct-marketing/code
https://github.com/Snowflake-Labs/snowpark-python-demos/tree/main/Credit%20Card%20Fraud%20DetectionLinks to an external site.https://github.com/Snowflake-Labs/snowpark-python-demos/tree/main/Retail-Churn-AnalyticsLinks to an external site.https://github.com/Snowflake-Labs/snowpark-python-demos/tree/main/Advertising-Spend-ROI-Prediction