Automated Data Analysis Framework

Development of a framework to load bank customer data and answer analytical questions using SQL and natural language queries for efficient and accessible data analysis.

Uday Lunawat

Problem Statement Overview

Challenges in Automated Data Analysis for Banking

Automated data analysis framework

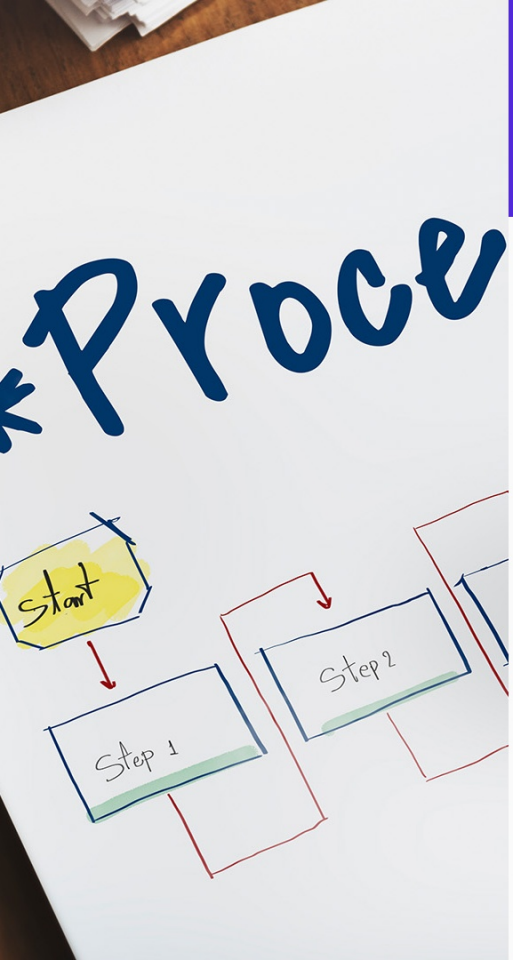
The need for efficient data analysis tools has grown, particularly in banking sectors dealing with vast datasets.

Natural language processing

The framework should convert analytical questions in natural language into SQL queries for effective data retrieval.

Loading bank customer data

It focuses on loading bank customer datasets, essential for analyzing queries about customer behavior and trends.



Dataset & Data Loading

Standardizing banking customer data for analysis

Loading data from sources

The code loads a CSV or Excel file containing comprehensive banking customer data.

Column name standardization

It standardizes the column names to ensure consistency and ease of access for subsequent analysis.

Storing in a structured format

The data is organized in a format of a list of dictionaries, making it accessible for further use in tools.

Natural Language to SQL Conversion

How the Gemini 2.5 Flash agent translates natural queries into executable SQL commands

Agent uses Gemini 2.5 Flash model

The agent is configured with the Gemini 2.5 Flash model to process natural language queries accurately.

Examples of query types handled

Includes average balance by gender and job, and regions with highest job classification concentrations.

Interprets natural language queries

The agent interprets user queries phrased in natural language and understands their intent for SQL conversion.

Converts queries into SQL

Queries are converted into executable SQL statements to retrieve precise data insights.

Execution of SQL Queries

Utilizing pandas and pandasql for seamless execution

Seamless SQL Query Execution

The framework automates SQL query execution without user intervention, improving efficiency.

Integration with pandasql

Leveraging the pandasql library enables the use of SQL queries directly within pandas data structures.

Supports Special Commands

Handles special SQL commands, such as table introspection, enhancing flexibility for users.

User-Friendly Interface

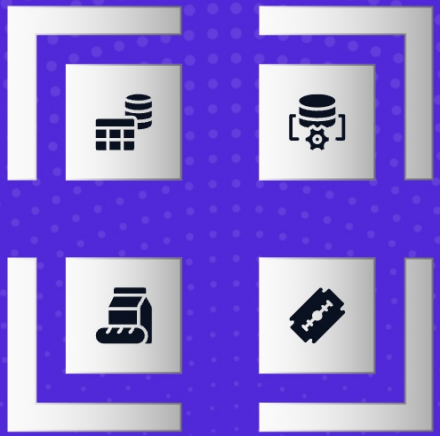
Users can execute complex SQL queries without needing to manage raw SQL syntax directly.

Integration & Modularity

Modular Design for Flexibility

Modular data preloading

The design incorporates a modular system for data preloading, promoting adaptability for varying datasets.



Flexible SQL execution

Each SQL execution tool can be invoked separately, enabling tailored queries based on user needs.

Adaptation for industries

The framework is designed to be adaptable for similar datasets in various industries, showcasing its versatility.

Easy framework extension

This modular approach allows for expansion with additional functionalities to meet future demands.

Use Case: Average Balance Analysis

Analyzing average balance by gender, age group, and job classification to understand customer segments

Understanding customer financial behavior

Insights gained from averages guide targeted banking strategies for various segments.

Grouping data by multiple dimensions

This query groups data by gender, age group, and job classification to segment customers.

Computing average balance values

Calculating average balances helps reveal financial behavior patterns across segments.

Use Case: Geographic Segmentation

Analyzing Job Classifications by Region

Identify Regional Concentrations

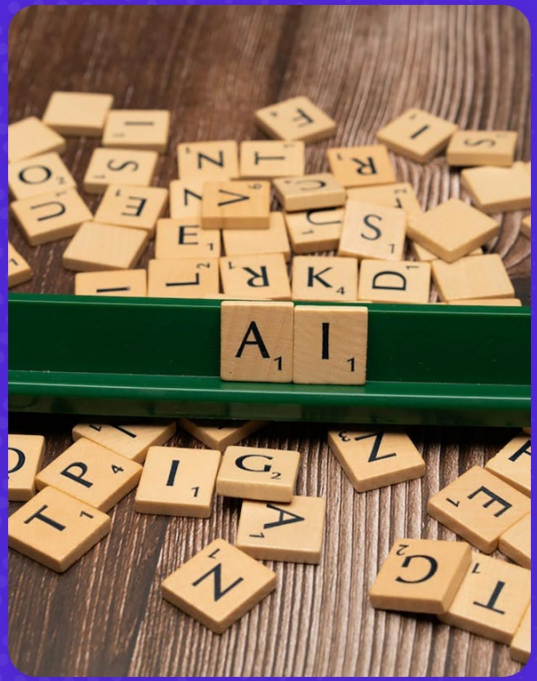
Highlight regions with a significant presence of specific job roles or classifications, enhancing targeted analysis.

Understand Employment Trends

Utilize geographic segmentation to observe how job roles shift over different areas, informing strategic decisions.

Overall Framework Design

Designing a system transforming language queries into data insights efficiently



Rapid transformation of language queries

System quickly converts natural language queries into actionable data insights for timely analytics.

Modular and scalable system design

Architecture supports complex, large-scale datasets with flexibility for future expansions.

Minimal manual intervention enabled

Automation reduces the need for manual input, streamlining data analysis workflows.

Targeted for global CPG clients

Designed specifically to address data challenges faced by global Consumer Packaged Goods companies.

Enhancing Insight Generation

Consider adopting this automated data analysis framework to enhance your organization's ability to derive insights from large datasets, ultimately leading to informed decision-making and improved customer understanding.