CS5224 Cloud Computing Preliminary Report

CostInsightSG

Team: PaaS the Cloud

Members: He Sirui A0297588Y, Hsieh Yu Hsuan A0304684B,
Guo Zhimao A0297939Y, Li Mengjie A0297851M, Shao Chen A0152077H

Table of Contents

Selected Theme

CostInsightSG - A Cloud-Based Platform for Cost of Living Analysis and Forecasting in Singapore

The cost of living in Singapore has long been a key concern for both policymakers and the general public. Macroeconomic indicators like the Consumer Price Index (CPI), personal income levels, and population demographics influence not only government decision-making but also corporate marketing strategies and personal financial planning.

However, there is currently no Singapore-centric platform that consolidates comprehensive data while offering visualization and predictive capabilities. CostInsightSG seeks to address this gap by integrating official datasets into a cloud-based solution that delivers interactive analysis and forecasting tools.

Business Model

The platform relies primarily on **advertisement** as its main revenue stream, providing free access to essential features to the public. This approach targets a broad initial user base, particularly students and private individuals who have a strong need for cost-of-living information but are less inclined to pay for it. By displaying ads, the platform can cover operational costs without charging for basic functionalities.

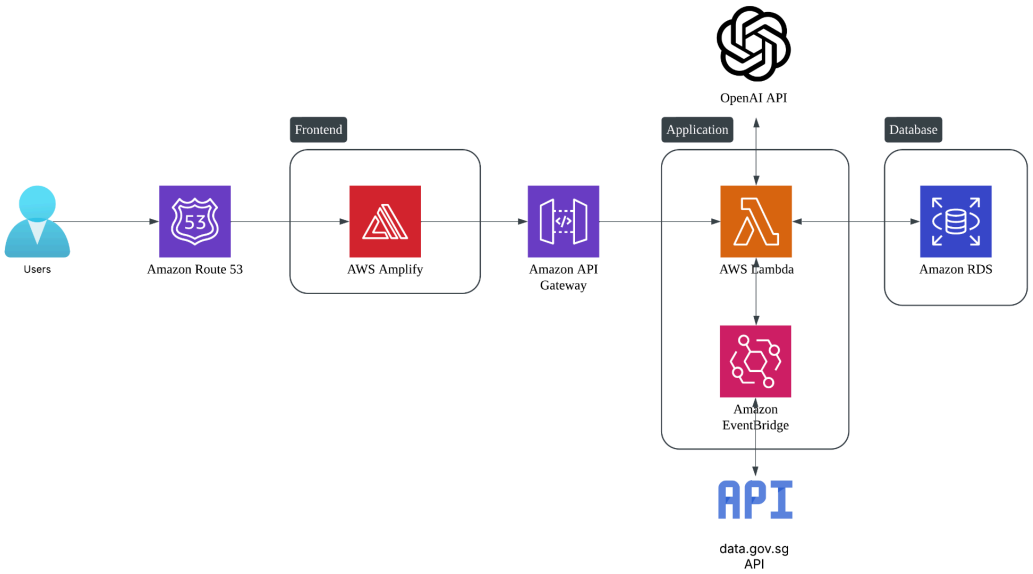
Cloud Service

CostInsightSG will be architected as a robust and scalable cloud-native platform, designed to democratize access to sophisticated cost of living analysis and predictive insights for individuals, businesses, and government agencies. By leveraging the inherent advantages of cloud deployment, including elastic scalability, high availability through redundant infrastructure, and a flexible pay-as-you-go pricing model, CostInsightSG will ensure accessibility and cost-effectiveness for a diverse user base. The platform will integrate data from a multitude of authoritative sources, including official government statistics, economic indicators, and potentially, aggregated consumer spending data, to provide a holistic and accurate representation of cost of living trends.

The core functionalities of CostInsightSG will be centered around empowering users with actionable insights:

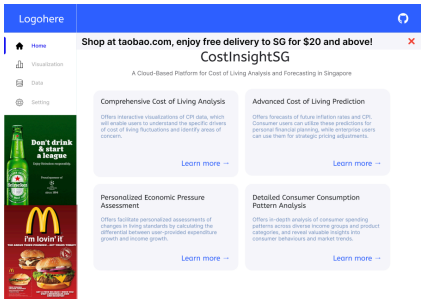
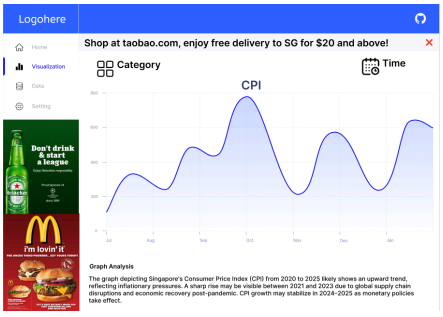
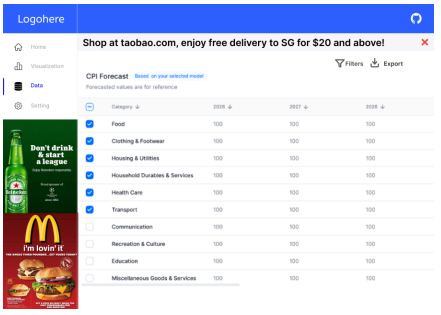
- Comprehensive Cost of Living Analysis: The platform will offer interactive visualizations of CPI data, including dynamic quarterly and annual trend charts. Users will be able to perform granular, deep-dive analyses across essential expenditure categories such as housing (rent, utilities), food (groceries, dining out), transportation (fuel, public transit), healthcare, and education. This will enable users to understand the specific drivers of cost of living fluctuations and identify areas of concern.
- LLM Model integration: For each visualization graph, a brief analysis paragraph will be generated using the OpenAI API, specifically the ChatGPT-4o model. To optimize token usage, a data table will store previously generated analyses. When a user makes a request, the system will first check the table for existing content. If the analysis is found, it will be retrieved from the table. If not, the system will call the OpenAI API to generate a new analysis, which will then be stored in the table for future reference.
- Advanced Cost of Living Prediction: CostInsightSG will incorporate a suite of sophisticated statistical and machine learning models, including time series analysis, and potentially, LSTMs, to generate forecasts of future inflation rates and CPI. Consumer users can leverage these predictions for informed personal financial planning, budgeting, and investment decisions, while enterprise users can utilize them for strategic pricing adjustments, inventory management, and long-term business planning.
- Personalized Economic Pressure Assessment: The platform will facilitate personalized assessments of changes in living standards by calculating the differential between user-provided expenditure growth and income growth. Users will have the flexibility to input income data manually or import it. By providing a clear picture of their financial health, this feature will empower users to take proactive steps to mitigate the impact of economic pressures.
- Detailed Consumer Consumption Pattern Analysis (Enterprise Focus): Targeted primarily at business and enterprise clients, this feature will offer in-depth analysis of consumer spending patterns across diverse income groups and product categories. By leveraging advanced data analytics techniques, including cohort analyses, the platform will reveal valuable insights into consumer behavior, market trends, and competitive dynamics. This will enable businesses to refine their market positioning, optimize product offerings, and develop targeted marketing campaigns.

Preliminary Design



Frontend

We will be using the Vue.js framework, together with Bootstrap and other components to build an interface to display cost of living data and analysis results, and AWS Amplify as a continuous integration and continuous deployment tool. Our page will create different modules based on the analysis content and provide interactive dashboards and data visualization services. In addition, we will use Amazon Route 53 to create a domain name and set the DNS to our front-end server IP address.

Homepage	Graph Representation	Table Representation
		

Backend

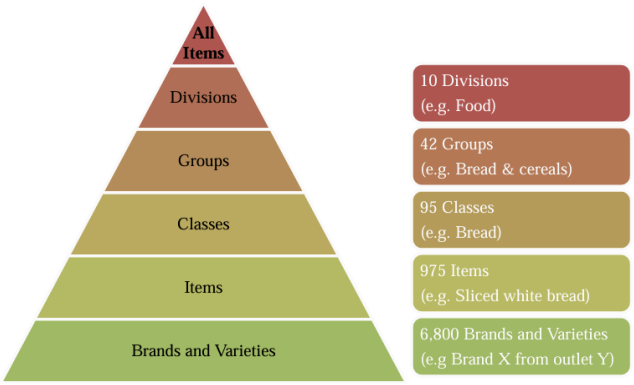
Since our entire SaaS platform will be hosted on AWS, the backend will utilize AWS RDS (Relational Database Service) for structured data storage and management. The database schema will be designed to accommodate multiple datasets, including but not limited to time-series consumer price index data, population indicators, and income distribution. Before being stored, the data will be pre-processed based on its type. Below is a list of potential data sources.

- Consumer Price Index (CPI), 2019 As Base Year, Monthly
- Consumer Price Index (CPI) By Household Income Group, Highest 20%, 2019 As Base Year, Annual
- Consumer Price Index (CPI) By Household Income Group, Middle 60%, 2019 As Base Year, Annual

- Consumer Price Index (CPI) By Household Income Group, Lowest 20%, 2019 As Base Year, Annual
- Income of Individuals by Income Type, Annual
- Cost Of Living Comparison

The data primarily comes from government sources (data.gov.sg), supplemented by cost-of-living comparisons from private websites. The latest data will be periodically fetched to ensure the service remains up to date, which can be accomplished using AWS Lambda and AWS EventBridge. Initially, data preprocessing will be handled manually. Once the core functionalities are successfully implemented, we will focus on developing an automated workflow for parsing the data. The data from data.gov.sg can be fetched with provided API, private cost-of-living comparisons will be extracted via web scraping. We'll also align data using 2019 as the baseline year to ensure consistency in comparison. A RESTful API will be developed using AWS API Gateway and AWS Lambda to allow frontend applications to query data efficiently.

AI Model Training



Data

The data that will be used for training and prediction are the 10 divisions (categories) from the dataset.

Data Processing

The first step involves visualizing the CPI time series to identify trends (e.g., long-term increases) and seasonality (e.g., recurring monthly patterns). Stationarity will be assessed using the Augmented Dickey-Fuller (ADF) test. If the data is non-stationary (p-value > 0.05), differencing will be applied until stationarity is achieved. Seasonality will be confirmed via autocorrelation analysis, expecting a 12-month cycle typical of monthly data. Missing values or outliers will be addressed through interpolation or statistical correction (e.g., Z-score).

Dataset Splitting

The dataset will be split chronologically:

- Train Set: January 2014 to December 2020 (84 months) for model fitting.
- Validation Set: January 2021 to December 2022 (24 months) for validation.
- Test Set: January 2023 to December 2024 (24 months) for testing.

Model Selection

Primary choice

Given the time series nature of CPI data, a Seasonal ARIMA (SARIMA) model will be chosen due to its ability to capture trends and seasonality. SARIMA has 2 components:

- Non-seasonal (p, d, q): Handles trends and short-term dependencies.
- Seasonal (P, D, Q, s): Captures repeating patterns over a fixed period.

Optimal parameters will be selected using Akaike Information Criterion (AIC).

Alternatives

Holt-Winters exponential smoothing or LSTM could be explored, though SARIMA is prioritized for its efficiency with 132 months of data. If category-specific CPI data is leveraged, an ARIMAX model could incorporate these as exogenous variables.

Prediction Target

The primary goal is to forecast the overall CPI for the next 12 months (January to December 2025), providing insights into future inflation trends. Model performance will be evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) on test data.

Limitation

Data Accuracy and Availability: The quality and timeliness of the analysis heavily rely on the accuracy and availability of underlying data sources (CPI, income, expenditure, etc.). Delays or inaccuracies in government data releases would directly impact the platform's reliability.

Data Integration Challenges: Integrating data from diverse sources with varying formats and frequencies can be complex and time-consuming for maintenance. Ensuring data consistency and avoiding redundancies can be a major hurdle.

Model Accuracy: Predictive models are inherently limited by their assumptions and the quality of training data. Economic forecasting is notoriously difficult, and unforeseen events (e.g., pandemics, geopolitical crises) can significantly impact accuracy. Overfitting models to a limited amount of historical data could lead to poor performance in predicting future trends.

Implementation Plan

The implementation plan and milestones are illustrated in the Gantt chart below.