

Midterm Project

Advertising Insights Project: Data-Driven Strategies for Global Campaign Optimization

Project Overview

Imagine you are a data engineer tasked by a Global Advertising Director of a multinational company to prepare a comprehensive report for optimizing advertising strategies. The Director's key objectives are:

1. **Identify regions and topics with high advertising potential** to focus on key markets.
2. **Analyze ad performance** to identify the best-performing ads and areas for improvement.
3. **Build predictive models** to improve targeting and conversions.
4. **Present findings in a professional and actionable format** to guide global strategy.

You will utilize data integration, analytics, and machine learning to accomplish these objectives by leveraging Google Cloud Platform (GCP) APIs through Python, including the BigQuery API, Cloud Storage API, and BigQueryML API. The project consists of several tasks to create an actionable business report.

Datasets

1. Google Trends Data

- Source: `bigquery-public-data.google_trends.international_top_rising_terms`
- Description: Insights into regional search trends and popular topics.
- Key Fields:
 - `country_code`: Region name code.
 - `country_name`: Region name.
 - `term`: Search term.
 - `score`: Interest score.

2. Google Analytics Sample Dataset

- Source: `bigquery-public-data.google_analytics_sample.ga_sessions_20170801`
- Description: Website user activity data, including page views and transactions.
- Key Fields:
 - `geoNetwork.country`: User country.

- totals.pageviews: Total page views.
- totals.transactions: Total transactions.

3. Google Ads Geotargets Data

- Source: bigquery-public-data.google_ads.geotargets
- Description: Information about geographic regions targeted by ads.
- Key Fields:
 - criteria_id : Geographic target ID.
 - en_name : Geographic target name.
 - country_code: Country code.

4. Google Ads Geo Mapping Data

- Source: bigquery-public-data.google_ads_geo_mapping_us.ads_geo_criteria_mapping
- Description: Mapping data linking geographic IDs to regions.
- Key Fields:
 - ads_criteria_id: Geographic ID.
 - target_city: Geographic location.

5. Dataset 1: Customer Interaction Logs

- Source: CSV uploaded to Cloud Storage (customer_logs.csv) via Pythion
- Description: User interactions with ads.
- Key Fields:
 - log_id: Log ID.
 - timestamp: Interaction time.
 - ad_id: Ad ID.
 - user_action: User action (click, view, purchase).
 - country: Geographic location.

6. Dataset 2: Ad Performance Data

- Source: CSV uploaded to Cloud Storage (ad_performance_logs) via Pythion
 - Description: Detailed performance metrics for ads by region.
 - Key Fields:
 - ad_id: Ad ID.
 - Nested Field of ad_performance:
 - region_id: Region ID.
 - impressions: Ad impressions.
 - clicks: Ad clicks.
 - conversions: Ad conversions.
 - country: Geographic location.
-

Tasks

Task 1: Data Preparation and Integration

Objective: Integrate datasets into a unified analytics table for further analysis.

Steps:

1. Ingest Data:

- Use Python APIs to extract and upload:
 - BigQuery API: Google Trends, Google Ads Geotargets, Geo Mapping, and Google Analytics datasets.
 - Cloud Storage API: Upload customer_logs.csv & ad_performance_logs.csv and link as the external BigQuery tables

2. Transform Data:

- Normalize 'score' in Google Trends to a [0, 1] range.

3. Integrate Data: at least one-thousandth (1/1000) of the data be randomly selected for analysis.

- Join datasets:
 - Match Ads Geotargets and Geo Mapping on criteria_id.
 - Combine Trends data with Ads Geotargets using country_code.
 - Merge all datasets using country.
-

Task 2: SQL Analysis

Objective: Use SQL to extract actionable insights.

Steps:

1. Regional Trends:

- Rank term within each country by score using RANK().
- Find out the top 5 terms per country.

2. Correlations:

- Calculate correlations between score, conversions, page views, and transactions for each country

3. Ad Performance:

- Compute the average click-through rate (CTR = clicks / impressions) for each ad_id and country_name.

4. Statistical Approximations:

- Use APPROX_COUNT_DISTINCT() to estimate unique country_name.
-

Task 3: Machine Learning with BigQueryML

Objective: Predict ad conversions using a linear regression model.

Steps:

1. **Feature Selection:**

- Extract features: score, CTR, user_action (=click)

2. **Model Training:**

- Train a LINEAR_REG model to predict conversions.
- FROM integrated_ad_analysis;

3. **Model Evaluation:**

- Evaluate MAE, RMSE, and R² scores.
-

Task 4: Dashboard Creation

Objective: Visualize insights and present recommendations using Looker Studio.

Steps:

1. **Dashboard Design:**

➤ **Page 1: Regional Insights**

- Heatmap: CTR by region.
- Bar chart: Top term count per country.

➤ **Page 2: Ad Performance**

- Line chart: CTR trends over time.
- Bar Chart with Error Bars (Absolute Error): Create a combined bar chart representing each ad's predicted and actual values, x-axis represents ad_id; within each ad_id, individual bars represent the predicted and actual values of conversions.

2. **Advanced Features:**

- Add filters for country, ad_id, and term.
 - Enable drill-downs for detailed analysis.
-

Report Writing Guidelines

The final report should be addressed to the **Global Advertising Director** and structured as follows:

1. **Introduction (5%):**

- Briefly explain the objectives of the analysis.
- Describe the datasets and their relevance.

2. **Data Preparation and Integration (10%):**

- Summarize the data ingestion and transformation process.
 - Include the schema of the unified table 'integrated_ad_analysis.'
3. **Key Insights (25%):**
- Highlight the findings/insights from SQL analysis.
 - If the current analysis doesn't yield insights, are there other analyses you can conduct that might uncover business insights? **(Bonus 5%)**
4. **Predictive Models (25%):**
- Explain the purpose and results of the conversion prediction
 - Include model evaluation metrics. How is the model performance?
 - If the original model performed poorly, how did you improve it? Please provide an updated version of the model.
5. **Dashboard Summary (15%):**
- Describe the dashboard's structure and how it supports strategic decision-making.
 - Include annotated screenshots.
 - If the current visualization doesn't yield insights, are there other visualizations you can conduct that might uncover business insights? **(Bonus 5%)**
 - Find the business insights for three countries
6. **Conclusion and Recommendations (20%):**
- Provide actionable recommendations based on analysis and models (e.g., identifying underperforming regions and suggesting increasing ad spending for top-performing terms.)
 - Suggest next steps for future data collection and analysis.
-

Submission Instructions

Please follow these instructions carefully to ensure your midterm project is submitted correctly:

1. Report Submission:

- Submit your report as a PDF file.
- The report's first page must include your name and student ID.
- The second page must include screenshots showing that your project has been successfully closed and the GCP billing page indicating the cost breakdown for the midterm project. **Reports missing this information will not be graded and will result in a score of zero.**
- The report content should adhere strictly to the six-point guideline provided above.

2. File Naming and Submission Deadline:

- Submit your report and Python files (named using your student ID) via MOODLE by December 11, 2024, at 11:59 PM. **Late submissions will not be accepted.**

3. Collaboration Policy:

Group discussion is allowed, but all submitted reports must be original. Plagiarism (the level of similarity is over 50%) will result in a score of zero for all involved parties.