| TIGER ANALYTICS |
| --- |
| Springboard Project: Emerging Business Opportunities |
|  |
|  |
| **Karthik Ragunathan, Vignesh Vijayarajan** |

|  |
| --- |

# Background:



Our client (Manufacturer A) is a leading Food & Beverage manufacturer. Client wants to understand the growth patterns of consumer preferences (themes) and evaluate positioning of their brand across different themes. Client also wants to know the sales drivers of their products.

# Data:

Client will provide the following data for the project:

* Sales Data – At UPC level for both Client and Competitors
* Social Media Data – Mentions of theme across all Social media Platforms
* Google Search Data – Search volume of the Theme
* Theme\_Product\_List – Product to theme Mapping
* Product\_Manufacturer\_List – Product to Manufacturer Mapping
* Theme List – Theme Names

Company will also provide a clear and detailed description of the dataset including all the fields present. Company will make available knowledgeable personnel to provide any necessary background on data and business context. These personnel will also help Tiger in working with the end consumers of our models and identifying which features could be actionable. They would also help in corroborating findings from the models and can help provide an insider’s business perspective

**Deliverables:**

**Note: The case has to be solved using** [**Code Templates**](https://sites.google.com/a/tigeranalytics.com/knowledge-tiger/IP-assets/Code-Templates)

Data Preparation from social media, google search and sales data.

Demonstrates:

* Provide the list of themes available across all data sources
* Understands consumer preference(themes) available in each data source
* Provide a report for data sufficiency, sparsity and anomalies in each data source
* Recommend the time granularity (Daily/Weekly/Monthly/Quarterly/Yearly) for the analysis

Data exploration and Hypothesis Validation

Demonstrates:

* Merge the required data sources
* Understand the overall market share of our client
* Find the potential competitors for our client in each theme
* What are the themes which are emerging in social media, Google Search & Sales?
* Validate the hypothesis: Trend flows from Social -> Search -> Sales
  + What is the latency observed?
  + Is the latency significantly different across themes?
  + Pictorially represent transition between sources

Build the sales model and identify the driver of sales

Demonstrates:

* Perform appropriate data transformation/aggregation
* Create a dependent variable by aggregating sales of our client to corresponding theme level
* Identify the right model technique and select the suitable variables
* Estimate the impact on sales due to social trends, search trends, own price and competitor effects
* Justify the estimated impacts are accurate
  + Model Performance
  + Hold out validation

Recommend levers for business growth

Demonstrates:

* Using EDA and Sales models, identify the themes with high business opportunity
* What are the controllable factors which the client could leverage to increase sales across themes?
* How to achieve a 5% increase in sales overall?

## 

## Code Templates (CT)

* You have to use ***Regression Code Templates (CT)***
* Use CT as a starting point and make customizations based on your solution
* Share your final codes along with EDA and modeling reports generated by CT with mentors ahead of the presentation date through github repo

## Deliverable Best Practices

* Structured code base with a few tests
* The submitted code should satisfy coding standard and the `QC` tests should all pass
* Documentation for the project using Sphinx (use case overview, modeling approach, evaluation reports etc)
* The README should contain, or point to a doc that contains, information about the dataset and how to run the notebooks/code.
* Notebooks should only have relevant analysis and should run successfully in a sequential manner. Notebooks should have adequate documentation on the analysis using proper markdown.
* Prepare the presentation in such a way to be able to cover it within 25-30 mins, excluding Q&A. Total number of slides should not exceed 25
* Presentation should cover the following sections: objective, exploratory Analysis, model results, final conclusion.
* Presentation should cover the answers of all the questions asked in the problem statement.

## Aspects considered during code evaluation

| **Module** | **Best practices to be followed** |
| --- | --- |
| **Submission Format** | - Follow the standard submission format  - Organize reports properly |
| **Code Quality** | - Configure data catalog properly  - Maintain neat and clean codebase  - Use proper variable names  - Make sure notebooks are running without any error  - Make sure notebooks are clean and sections are clearly organized |
| **Data Discovery** | - Identify all the datatypes properly  - Perform data validation  - Add/modify data rules, remove unnecessary rules |
| **Data Preparation** | - Merge/concatenate all the data sources together using appropriate keys  - Prepare the data as per the desired granularity, without any duplicates  - Perform pre-post merge checks  - Save intermediate data (cleaned, processed etc.)  - Perform train/test split properly |
| **EDA** | - Have a clear understanding of target and independent variables.  - Generate Univariate and Bivariate Reports as shown in reference notebooks  - Take actions based on the understanding from the reports  - Describe decisions influenced by EDA |
| **Feature Engineering & Transformation** | - Perform data transformation as per the guidelines  - Perform missing values, outlier handling  - Create reusable components for data transformation  - Follow standard hygiene and practice for data transformation  - Create new features to improve the model performance |
| **Feature Selection** | - Remove irrelevant features beforehand  - Reduce multicollinearity for the linear models  - Use techniques to identify important features  - Perform feature selection with business sense |
| **Model Evaluation, Comparison & Final Model Selection** | - Use relevant tools to evaluate and compare models  - Explain model outcome using interpretation module in reports  - Use techniques to remove bias from model  - Perform multiple model iterations based on the understanding from the reports |