|  |
| --- |
| 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:**



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 client could leverage to increase sales across themes?
* How to achieve a 5% increase in sales overall?

## 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.
* Please make use of the ***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
* 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