****

**ETICC support**:

ETICC offers a feasibility study of the existing &Facts system. The purpose is to understand the existing system and how to improve on it.

**Architecture**:



**Goal 1 - Data acquisition**:

Study how to retrieve, integrate and process market data from different data sources and any types of channels: marketing, transactions, logs, APIs, Web scrapping. Open data often comes in the form of market reports, industry insights and online surveys;

**Why gather data**:

With the aid of data, SMEs can go global more speedily and at a younger age. Big Data is a valuable asset that can be used to predict customer buying habits, detect fraud and waste, analyze product sentiment, and react quickly to events and changes in business conditions. It can also help discover new business opportunities. The data can be gathered from social network activity, web server logs, sensors used to gather climate information, stock market data, e-mails, transaction records, web click streams, etc. data coming from geolocation and social media are more important and more often exploited by SMEs. Global data as a service market is growing increasing demand for statistical analysis. The organizations nowadays tend to outsource the big data services to other companies. Data modeling is required, in which raw data is collected, cleansed, categorized, converted, aggregated, validated, and otherwise transformed. Once the data is clean, it's stored in a structure and format that lends itself to reporting.

**Goal 2 - Data integration with ETL (Extract, Transform, Load)**:

Hadoop provides scalability and flexibility out of the box. The raw data needs to be preprocessed from the initial sources and sent to a warehouse. The steps are:

1. Retrieve data from the data sources;
2. All extracted data is analyzed and the duplicates are removed. The result is then standardized, sorted, filtered, and verified;
3. The transformed data is uploaded into the warehouse. The Data Warehouse (or RDBMS) is used to do the Queries.

Apache Hadoop takes care of the ETL and the distributed processing through:

1. HDFS, a batch processing engine stores the data in a distributed manner;
2. MapReduce divides the task submitted by the user into a number of independent subtasks and executes them in parallel;
3. YARN is a resource management layer that provides resource management for Hadoop

**Data integration**:

+ Middleware data integration: layer between two disparate systems allowing them to communicate;

+ Data Virtualization Integration Approach: leave data in the source systems while allowing to create a new set of united views;

+ Data Warehouse Approach: database which stores a unified version of data extracted from all the source systems involved and manage it independent of the original source systems. manage history of data (or data versioning). combine data from very disparate sources (mainframes, databases flat files, etc.) and to store them in a central repository of data;

+ Semantic Web Approach.

**Data integration challenges**:

The challenges are uncertainty of data, management, syncing across data sources, finding insights, and skill availability. Data integration is managed by Extract, Transform and Load (ETL) tools such as Data Warehouses, but these techniques are now inefficient. The primary obstacles are:

+ Amount of data;

+ Storing and managing the big data;

+ Integrating variety of data and extract knowledge;

+ Inability to utilize the data without wasting it;

+ Costs.

**Goal 3 - BI Tool for Analyzing Data**:

Use AI and ML to develop a business intelligence tool. Data Analysis is a process of collecting, transforming, cleaning, and modelling data with the goal of discovering the required information.

**Business Intelligence**:

Business intelligence (BI) transforms data into insights that inform an organization's business decisions. It presents those findings in reports, summaries, dashboards, graphs, charts and maps to provide users with detailed intelligence about the state of the business. This is what is called descriptive analytics. Descriptive analytics uses a full range of data to give an accurate picture of what has happened in a business and how that differs from other comparable periods. These performance metrics can be used to flag areas of strength and weakness to inform management strategies.

**Types of Analytics for BI**:

Data Analytics and visualization metrics are different types of statistical measures, such as:

+ Web Analytics: collection, reporting, and analysis of website data;

+ Market Trends Analysis: an analysis of past and current market behavior and dominant patterns;

+ Market Segmentation: Segmentation allows determining similar customer groups based on data, allowing for a broader picture of potential customers;

+ Customer Analytics: data from customer behavior is used to help make key business decisions via market segmentation and predictive analytics. Behavioral profile (why your product matches a customer's lifestyle) and a demographic prole (describing a customer's demographic attributes);

+ Customer Churn Prediction: which customers are likely to cancel a subscription to a service based on their use of the service;

+ Geospatial Data Analytics: how e.g., sales data relates to other data;

+ Competitor Analytics: pick the right competitors to analyze and Know which aspects;

+ Supply Chain Analytics: aims to improve operational efficiency, by enabling data-driven decisions at strategic, operational and tactical levels;

+ Digital Marketing Analytics: view of what marketing strategy is working;

+ Pricing Analytics: how pricing activities affect the overall business;

+ Social Media Analytics: data from social media sites and blogs and evaluating that data to make business decisions;

+ Performance and Strategy Management: provide insight to guide and support the performance and strategy management processes of a business and track business performance versus objective.

**Goal 4 - BI Tool for Reporting and Mining (Data visualization dashboard)**:

AI-powered reporting backend formulates key insights and generates data visualizations such as tables, graphs, and charts. Data Mining is a process of extracting the rules from large quantities of data.

**Goal 5 - Configure an Enterprise Data Warehouse**:

This is used to store, process and query existing research data. It is a large centralized repository of data that transform all multiple data formats into a single format and consolidating them in one place. data mart is a subset of a data warehouse oriented to a specific business line. Data marts contain repositories of summarized data collected for analysis on a specific section or unit within an organization. warehouse is a database that keeps information in a predefined format. A warehouse is connected with data sources and ETL systems on one end and BI tools or dashboard interfaces on the other. This allows for presenting data from various systems via a single interface.

**Goal 6 - Develop a stronger &Facts Proof of Concept or Minimum Viable Product**:



It is desirable to implement a Proof-of-concept application for the company with all the required requirements. The requirements of the application are + Allow secure access to the system;

+ Allow authorized users to manage (add, change, delete) user rights;

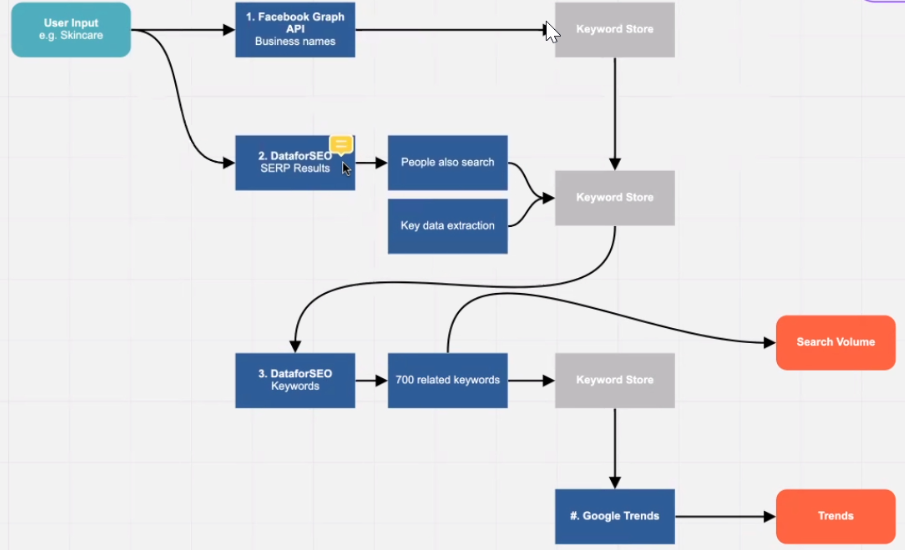
+ Only authorized user can perform search and visualize their market trends;

+ Import and integrate data from different sources;

+ Provide notifications to profile-holders;

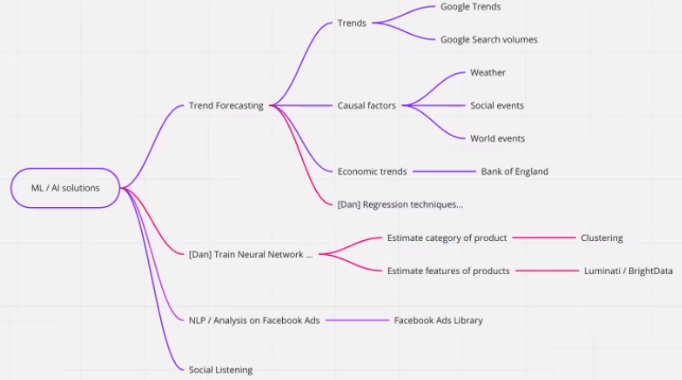
+ Take payment from the user.

**Keywords Generation**:



**Use Cases**:

+ ‘Skincare’ (or other) is fed to Facebook Graph API and DataforSEO to generate some keywords. DataforSEO gives a list of ‘what other people search’. From this a list of keywords can be extracted and used again to search DataforSEO to generate 700 other related keywords. On this keywords DataforSEO returns the Search Volume, that is how much people are searching for this data. When enough keywords are generated, they can be fed into Google Trends which returns a list of trends.



+ “Trend Forecasting/Trends” (Compulsory): the trends and ‘search volumes’ from the previous search and also Luminati are used to first visualize and eventually perform forecasting. The trends and forecasts are visualized in a graph like the one at <https://www.retentionx.com/>, allowing the user to select the dates and forecast method. Forecasts can be performed using geometric models, S-Curves and/or other models (Dan Regression)



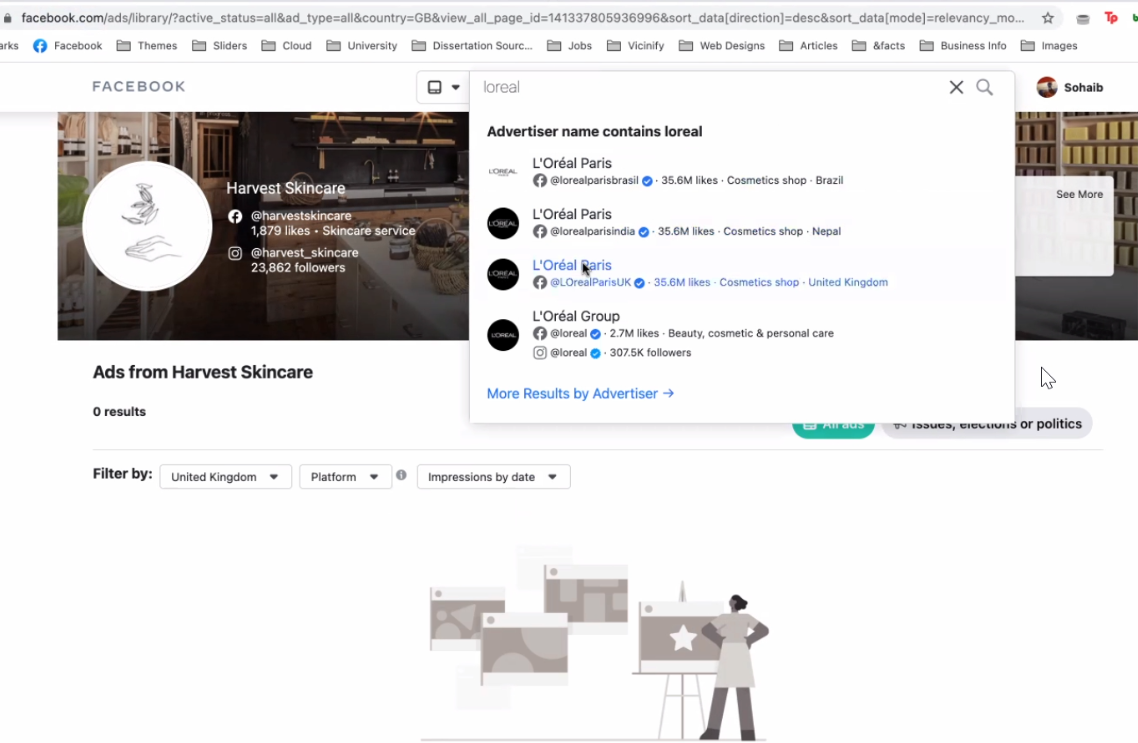
The graphs can be implemented in the Front end for example using the framework called ‘minimals’ <https://minimals.cc/docs/typography> which uses React. React is optional, Aurelia can be used instead.

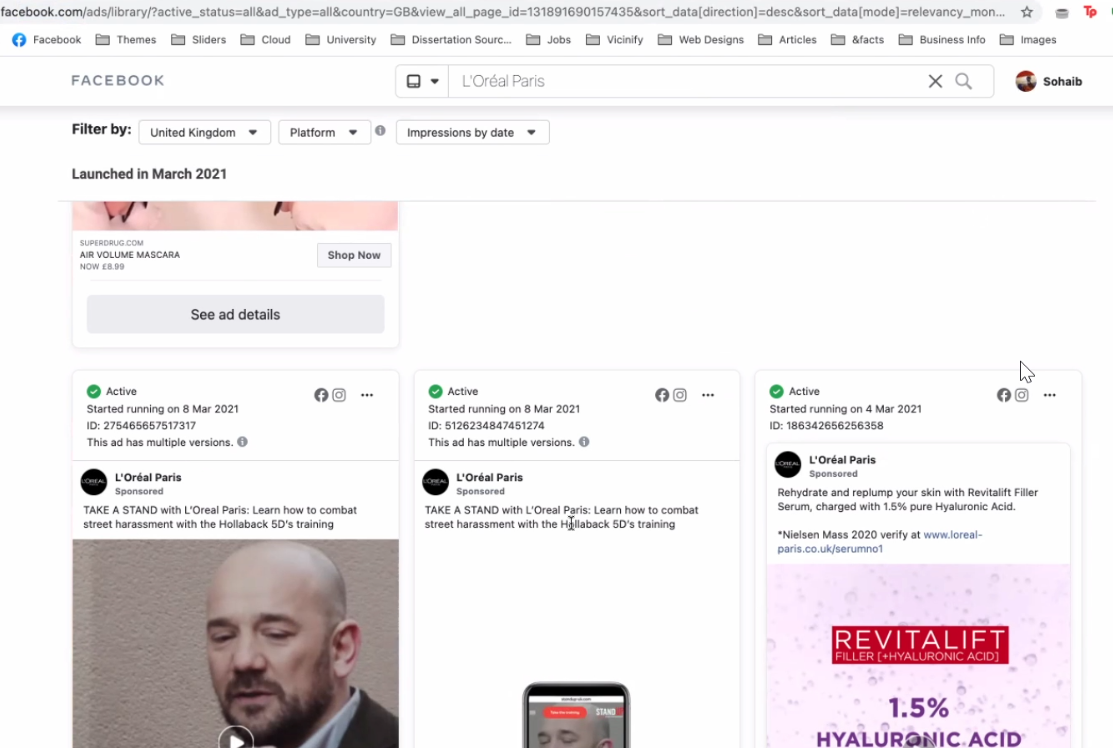
+ “Trend Forecasting/Causal Factors” (Compulsory): These are much a product or company or trend is affected by external circumstances. Causal factors won’t be from the Google trends API. They will be through different APIs or data sources, which are yet to be fully identified;

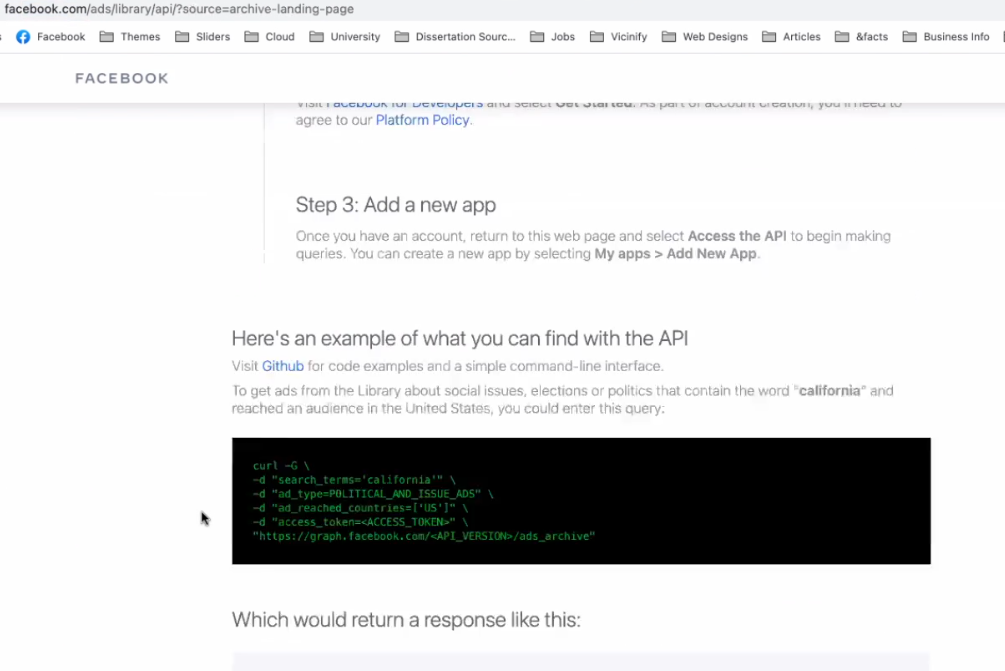
+ “Trend Forecasting/Economic trends” (Compulsory): This is how much money the company or product is making. I am not aware of a BoE API. We can either web scrape, or request large datasets from BoE - we might be able to set up a feed. I will have a chat with my contacts at BoE nearer the time;

+ “Train neural networks” (Compulsory): train neural networks using data in order to guess features. Build some features to create clusters. For example, one cluster might represent companies that would benefit from taking on bigger debt. Classify a company to ‘success/failure’ percentages. For example, a company might probably fail but it can be shown that if it applies the insights of AndFacts this will bring it to success;

+ “NLP Analysis on Facebook Ads” (Optional): The Facebook ads library returns a list of ads associated with a trend. The applications can suggest to a client what ads can work for their business based on what their competitors are doing.







+ “Social listening” (Optional): this uses the APIs of each social media to get data from users to understand the sentiment of users