**Service providers getting advantages with big data analytics**

With growing demands of an ever-progressive customer, retail needs to keep pace with evolving trends. With the global localization, new strategies need to be adopted by retailers in all economies to remain relevant in the customers mind. Predictive Analytics can provide the framework within which intelligent decision making could be made possible with higher business efficiencies.

Service providers are sitting on terabytes of data that is stored in siloes and scattered across the organization, to exploit the full potential of that data, service providers must have solutions that can help them correlate, process and decipher nuggets of actionable information. This is only possible through the big data and advanced analytics.

I feel big data used effectively has the potential to revolutionize the way telecom operators build, run and market their services. In this project I am trying to use advanced analytics for more innovative business models that can offer more targeted campaigns based on customer segmentations, as well as location and I am trying to concentrate more on increasing sales, real time analytics and decision making by collecting live data sets from AT&T organization and analyzing those huge data sets for the above-mentioned goals. I am mostly emphasizing on creating solutions that can combine customer usage and subscription data with insight customer mood and preference and network availability, to trigger specific actions. For this analysis I would like to use Clustering (Apriori) algorithm and on top of it I would like to use streaming analysis using python scripting and on the other side I would like to pull reports with different data combinations using Tableau tool.

**Notable aims of the project**:

1. Based on previous data of the region, I would like recommend a service type to a particular region where there is already a related service type exists.
2. Based on previous data of the region, I would like recommend a new plan type to a particular region where there is already a related plan exists.
3. Recommending new plans and service types by comparing different service provider’s plans and their region of implementation.

**Related domain of study**: Retail industry (Telecom service providers-AT&T)

**Key steps and Algorithms to be used**:

1. Cleaning the data and creating Numpy’s and panda’s.
2. Data visualization.
3. Twitter streams.
4. Association rules.
5. Clustering (Apriori algorithm).

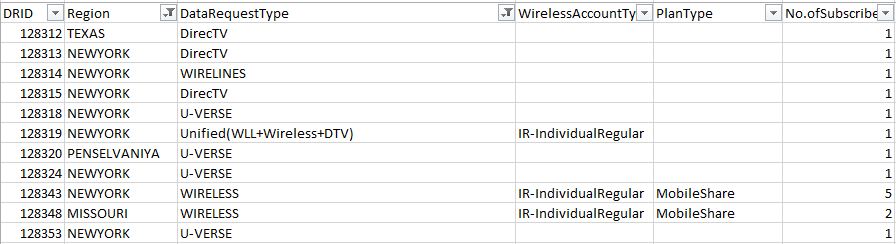
Before doing anything with the data sets, I have to clean them, this means removing all the unnecessary columns and fields from it. This makes data sets to be ready for analysis. Then I am trying to create numpy’s and pandas on the data sets so that I can play with the data and its numbers. From those mathematical calculations I am trying to derive some tables and graphs using data visualization for the core analysis of the data sets. With the help of the Association rules and Forecasting/prediction clustering (Apriori) algorithms, I am trying to answer the aims of this project. Apart from this, I am trying to scrape some data from social networking sites and do some analysis with that data to see the public response that helps a lot for recommending engines to an organization. I am also using tableau tool to implement all the above stated data in graphical format for better understanding on the analysis.

Of all the clustering algorithms I have chosen Apriori algorithm because Apriori is an algorithm for frequent item set mining and [association rule learning](https://en.wikipedia.org/wiki/Association_rule_learning) over transactional [databases](https://en.wikipedia.org/wiki/Databases). It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine [association rules](https://en.wikipedia.org/wiki/Association_rules) which highlight general trends in the [database](https://en.wikipedia.org/wiki/Database).

I am trying to interpret the results obtained and give some recommendations to get more profit by an organization by implementing some particular service type and particular plan in a suitable region.

**Data sources**: Live AT&T data sets.

I have taken the live data sets from AT&T, which contains the region, service type and its associated plan type in that region, with number of subscribers, I am giving a glance of that data here.



I will use pandas on this data set, so as to make it flexible for mathematical calculations and prepare it for data visualization input. I have to prepare this data for using Apriori algorithm as well to make the best out of it.

**Graphics:**

I am trying to represent the data bar graphs and pie charts for better visualization. With the help of the bar graphs we can understand the relation between region and service type and its various plans, and with the help of pie chart there will be a better understanding market contribution of the plans mentioned. Finally when we analyze data with Apriori algorithm we get the best graphs giving recommendations about the implementation of new service and plan types in a region.

**Current challenges:**

1. Hard to clean the live datasets from AT&T, and changing them into required format for Apriori algorithm.

2. Understanding and combining twitter data analysis with the recommendations from Apriori algorithm.

**Citations:**

<https://github.com/asaini/Apriori>

<http://aimotion.blogspot.com/2013/01/machine-learning-and-data-mining.html>

<http://blog.hackerearth.com/beginners-tutorial-apriori-algorithm-data-mining-r-implementation>

<http://adataanalyst.com/machine-learning/apriori-algorithm-python-3-0/>

<https://codereview.stackexchange.com/questions/136169/apriori-algorithm-in-python-2>

<http://pbpython.com/market-basket-analysis.html>

<https://www.tableau.com/learn/training>