|  |
| --- |
| Vehicle Demand Forecasting (VDMS)  Solution Design Approach |

**High Level Forecasting Process**

1. Data Lake Formation & Data Engineering
2. Primary Data Exploration
3. Clearing Issues, observations, Queries, pending items mentioned in attached Excel
4. Understanding / discussion on data - Primary key’s in various tables to understand uniqueness and foreign fields to join with tables.
5. Business Process mapping / Data Mapping to critical features for forecasting
6. Mapping of received data fields - with the target set of critical features for forecasting i.e. Sales Information, Cancellation Information, Enquiry, Promotions data, Exchange Data etc. as same information  in multiple tables  -
   1. Enquiry ( Enquiry Strength )
   2. Booking
   3. Invoicing
7. Data Exploration - with Selected key Features
8. Data Modelling –Multiple Iterations
   1. We used DeepAR ( AWS platform)
   2. We used Prophet - given by Facebook ( using Sage maker of AWS), used Python as a scripting lang.
   3. Also used HPO functionality of AWS platform – Hyper parameter optimization
9. Training the Model
10. Test Model / Model Accuracy
11. Comparison with Manual Forecasting

**For VDMS Demand Forecasting Solution – we considered following Data sets (Iteration -1)**

1. Sales Data
2. Enquiry Data
3. Seasonality Holidays
4. Competition Data
5. Macro-Economic Indicators

**Our Solution approach is multistep** –

1. Aggregate Invoice & Enquiry data on monthly basis to form time series of sales and enquiry respectively, did weekly for generating more data points

As there are 121 variants in the data but for many of the variants count of data points are not sufficient for forecasting, so we employed a criteria to choose best variants for which forecasting can be done.

* To increase data points we took weekly aggregation of data instead of monthly, if we take monthly we don’t get sufficient data points even for these 12 variants.
* We choose those for which data is available for previous year i.e. most recent data points are available for those variants.

1. Decompose time series into constituents to observe components like seasonality and trends
2. Check for various properties of the series e.g. stationarity etc. using statistical tests
3. did data Partitioning
4. Modelling the data for existing patterns using various statistical / Machine Learning model e.g. DeepAR, ARIMA, smoothing techniques etc.
5. Comparing the results for various models
6. Also, comparing with injecting other data competition and other economy indicators
7. Test Model / Model Accuracy
8. Choose one best model
9. Training the Model
10. Forecast with model
11. Compare it manual forecast

We applied the statistical forecasting methods on following combinations of data

1. Run 1 - Sales data only
2. Run 2 - Sales data + Holiday Seasonality
3. Run 3 - Sales data + Holiday Seasonality + Enquiries data + Competitors data + macro-economic factors
4. Run 4 - Model wise WA & SR – Sales + Seasonality data, for SM and WM sufficient data is not available, we did only for two of the model them for model wise forecasting

|  |  |  |
| --- | --- | --- |
|  | Swift | Wagon R |
| 1st Iteration | SR | WA |

Note: - Data considered at rollover variants level.

We employed multiple methods to forecasts and observed results in various cases, four of the model runs are selected as current benchmarks

**Iteration -2**

For VDMS Demand Forecasting Solution – we considered following Data sets

1. Sales Data
2. Seasonality Holidays
3. Competition Data
4. Leading Macro-Economic Indicators
5. Social Listening Data
6. Website hits
7. Brand Track – MIND Share Parameter
8. Discounts & Promotions Data

Leading Macro-economic Indicators

1. Fuel rates
2. Inflation - CPI
3. Stock Price
4. GDP
5. Bond yields (%)
6. Capacity utilization (%)
7. Bank credit growth (%)
8. RBI\_repo\_Rates

We applied the statistical forecasting methods on following combinations of data

Sales Data – Swift and Wagon R data considered.

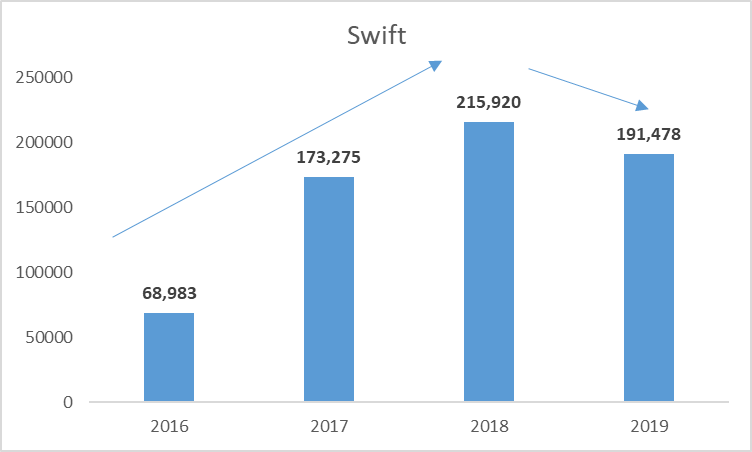
|  |  |  |
| --- | --- | --- |
|  | Swift | Wagon R |
| 2nd Iteration | SR+SM | WA+WM |

1. Model Run 1 - Sales Data + Holiday Seasonality + Websites Hits\_Swift WagonR
2. Model Run 2 - Sales Data + Holiday Seasonality + Social Listening
3. Model Run 3 – Sales Data + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share)
4. Model Run 4 – Sales Data + Holiday Seasonality + Discounts and Promotions Data
5. Model Run 5 – Sales Data + Holiday Seasonality + Discounts and Promotions Data + Websites Hits\_Swift WagonR + Social Listening + New Car Brand Track ( KPI – Mind Share)
6. Model Run 6 – Model Run 5 + Hyper Parameters Optimization
7. Model Run 7 – Sales Data
8. Model Run 8 – Sales Data + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Hyper Parameters Optimization
9. Model Run 9 – Sales Data + Holiday Seasonality + Discounts and Promotions Data + Websites Hits\_Swift WagonR + Social Listening + New Car Brand Track ( KPI – Mind Share) + Leading Economic Indicators + Competitor Data + Hyper Parameters Optimization
10. Model Run 10 – Sales Data + Holiday Seasonality + Discounts and Promotions Data + Websites Hits\_Swift WagonR + Social Listening + New Car Brand Track ( KPI – Mind Share) + Leading Economic Indicators + Competitor Data + Hyper Parameters Optimization (Combined Approach)

Sales Data

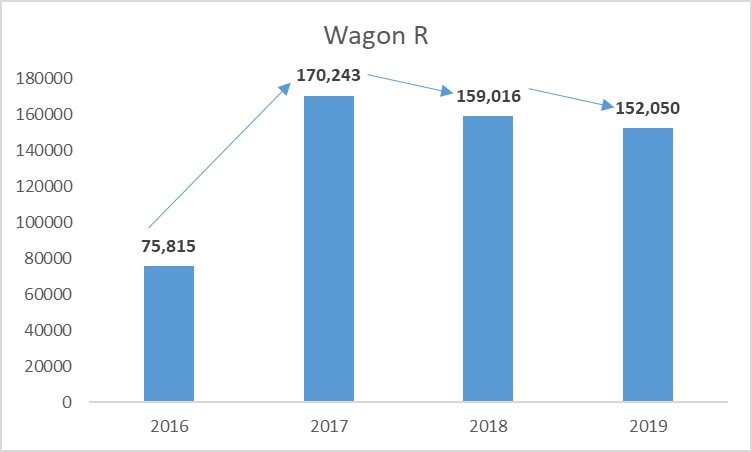
Trend Identification

|  |  |
| --- | --- |
| Data Summary |  |
| Count: | 174 |
| SUM(Units) |  |
| Sum: | 630,956 |
| Average: | 3626 |
| Minimum: | 257 |
| Maximum: | 11,977 |
| Median: | 3297 |
| Standard deviation: | 1839 |



Trend Identification

|  |  |
| --- | --- |
| Data Summary |  |
| Count: | 174 |
| SUM(Units) |  |
| Sum: | 535,724 |
| Average: | 3079 |
| Minimum: | 209 |
| Maximum: | 10,515 |
| Median: | 2898 |
| Standard deviation: | 1532 |



Trend Identification

After log transformation

**Learnings from 1st, 2nd and 2.1 Iterations**

* 1. Improve Models in 3rd Iteration by mapping daily data instead of weekly – and by trying more combinations and understand seasonality at more granular level
  2. Do weightage analysis & apply combinations accordingly for secondary data elements in a model running to identify best fit parameters
  3. Create month only model and observe the forecasts ( this did not worked)

When used weekly forecast and aggregated for a last month – it induced error for 3 days, which needed to be accounted for in final forecast.

Week ended was ignored so 2 days forecast ignores

**Iteration -3**

For VDMS Demand Forecasting Solution – we considered following Data sets for Jan-20 Forecast

1. Sales Data
2. Seasonality Holidays
3. Leading Macro-Economic Indicators
4. Social Listening Data
5. Website Hits - Showroom visits
6. Brand Track – MIND Share Parameter
7. Discounts & Promotions Data

Leading Macro-economic Indicators

1. Fuel rates
2. Inflation - CPI
3. Stock Price
4. GDP
5. Bond yields (%)
6. Capacity utilization (%)
7. Bank credit growth (%)
8. RBI\_repo\_Rates

We applied the statistical forecasting methods on following combinations of data

Sales Data – Swift and Wagon R data considered.

|  |  |  |
| --- | --- | --- |
|  | Swift | Wagon R |
| 3nd Iteration | SR+SM | WA+WM |

|  |  |
| --- | --- |
| **Model Runs** |  |
| [Model Run 1](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 1'!A1) | **Sales Data + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Discount+ Hyper Parameters Optimization** |
| [Model Run 2](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 2'!A1) | **Only Sales** |
| [Model Run 3](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 3'!A1) | **Sales Data + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Discount+ Hyper Parameters Optimization + December Data** |
| [Model Run 4](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 4 - Swift'!A1) | **Sales(Swift) + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Discount+ Hyper Parameters Optimization + December Data** |
| [Model Run 5](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 5 - WagonR'!A1) | **Sales(WagonR) + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Discount+ Hyper Parameters Optimization + December Data** |
| [Model Run 6](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 6'!A1) | **Sales Data + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Discount+ Hyper Parameters Optimization + December Data** |
| [Model Run 7](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 7 - Swift'!A1) | **Sales(Swift) + Bank Credit Growth + GDP + Hyper Parameters Optimization + December Data** |
| [Model Run 8](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 8 - WagonR'!A1) | **Sales(WagonR)+Hoidays+Mindshare+Social Listening Bond Yields+ Hyper Parameters Optimization + December Data** |
| [Model Run 9](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 9 - Swift'!A1) | **Sales(Swift) + Bank Credit Growth + GDP + Hyper Parameters Optimization + December Data** |
| [Model Run 10](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 10-WagonR'!A1) | **Sales(WagonR)+Hoidays+Mindshare+Social Listening+Bond Yields+ Hyper Parameters Optimization + December Data** |
| [Model Run 14](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 14 - Swift'!A1) | **Sales Data + Holiday Seasonality + Social Listening + Bank Credit Growth + RBI Repo Rate + GDP + Hyper Parameters Optimization + December Data** |
| [Model Run 15](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 15 - WagonR'!A1) | **Sales Data + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Website Hits +Showroom Visits + Hyper Parameters Optimization + December Data** |

|  |  |
| --- | --- |
| Appendix (with Refined New Model) | |
| [Model Run 11](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 11 - October'!A1) | **Sales Data + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Discount+ Hyper Parameters Optimization + September Data** |
| [Model Run 12](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 12 - November'!A1) | **Sales Data + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Discount+ Hyper Parameters Optimization + October Data** |
| [Model Run 13](file:///D:/Working/0%20MIRM%2019-20/AWS%20-%20Analytics/Maruti%20-%20Demand%20Forecasting/0.%20Final/Jan%20-%20Iteration%203%27/Sent%20to%20maruti%20-%2024Jan20/VDMS%20Jan-20%20Forecasts%203rd%20Iteration%20V1.0S(Part-1).xlsx#'Model 13- December'!A1) | **Sales Data + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Discount+ Hyper Parameters Optimization + November Data** |

Selected Optimized Models

|  |  |
| --- | --- |
| **Models** | **Description** |
| **Model 6** | **Sales Data + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Discount+ Hyper Parameters Optimization + December Data** |
| **Models Based on Weightage** | |
| **Model 14** | **Sales Data (Swift) + Holiday Seasonality + Social Listening + Bank Credit Growth + RBI Repo Rate + GDP + Hyper Parameters Optimization + December Data** |
| **Model 15** | **Sales Data (Wagon R) + Holiday Seasonality + New Car Brand Track ( KPI – Mind Share) + Website Hits +Showroom Visits + Hyper Parameters Optimization + December Data** |

Weight Analysis – we did weight Analysis of all the parameters in sales forecasting, and found few of the parameters are more useful in sales forecasting than the others – then we did forecasting using these parameters – Model 14 & Model 15 are created using such approaches only.

For Key Improvements we tried –

**Daily Data** - We tried to map seasonality in a best possible way and hence created models using Daily sales and tried with weekly sales also, As you know for statistical forecasting, it gives best results if there are sufficient data points as it has to clearly understand data patterns. So, to increase data points we took daily & weekly aggregation of data instead of monthly.

**Month Only Model** – we created monthly model for Jan only but not much success in this

**Parameter Weight Analysis**  – we did weight Analysis of all the parameters in sales forecasting, and found few of the parameters are more useful in sales forecasting than the others – then we did forecasting using these parameters – Model 14 & Model 15 are created using such approaches only.

**Learnings from 3rd Iterations**

* Improve Models by mapping daily data for longer durations – for about 8 or 9 yrs. As to cover all possible business cycles and thus improving forecast results further, as current business cycle is unique w’r’t to data used.
* Include more parameters i.e. natural calamity situations / flood etc. to refine the forecast
* Include harvesting calendar
* Do weightage analysis for inclusion of more parameters and then apply combinations accordingly for secondary data elements in a model running to identify best fit parameters
* Compute forecasts at state level which will further increase forecast accuracy

**Test Cases**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test Case No.** | **Description** | **Test Data** | **Expected Result** | **Actual Result** | **Status (pass/fail)** |
| #1 | Handling missing values | CSV file | Missing values checked | Missing values checked | Pass |
| #2 | Data integrity maintenance | CSV file | Data integrity maintained | Data integrity maintained | Pass |
| #3 | Trend stationarity | CSV file | Trend stationarity achieved | Trend stationarity achieved | Pass |
| #4 | RMSE value reduction check | CSV file | RMSE value  reduced | RMSE value  reduced | Pass |
| #5 | Model Accuracy | CSV file | Model accuracy 95% was expected | Model accuracy 93% achieved | Pass |

Solution Summary - Training & Handover Notes

**About Challenge(s)**

Maruti wanted to save costs by optimizing their sales forecasting to ensure right production plan. The challenge was to accurately forecast the sales of selected vehicle models for the upcoming months based on several time series datasets. Based on the forecast predictions they would adjust the production plan of selected vehicles. They previously relied on manual forecasting techniques with sub-optimal results.

Based on discussions with the Line of Business owners, the following key points were identified to address this with a ML solution.

* To forecast the sales of the selected vehicles models for the upcoming months based on multiple time-based economic indicators.
* Multiple data sources to be consolidated to get a holistic view. Some challenges with the data were
  + Data in multiple formats and granularity level
  + Missing data points
  + Public data with no defined interfaces

.

**Proposed Solution**

Based on discussions with customer, MIND team examined various data sources and data sets which can help to address the business problem, and proposed a ML based demand forecasting models, comprising of following steps

* MIND received three years data for selected vehicles to model it for for forecasting.
* We had scrapped economic indicators, competition, and holiday’s data from Internet. These economic indicators include Bank Credit Growth, Bond Yields, Capacity Utilization, RBI Repo rate, and Fuel rates, Stocks, CPI and GDP.
* Client also provided other secondary parameters like Social Listening, Brand Track, Website Hits, Discounts & promotions for modelling of demand.
* Did analysis on the data using multiple techniques, like exploratory data analysis, feature selection, dimensionality reduction (PCA), regression analysis. Did PCA for dimensionality reduction in order increase accuracy.
* Created ML models for demand forecasting –
  + - Tried Holts Winter, Deep AR models to map the demand
    - Created LSTM models to forecast sales of the selected vehicles using custom LSTM algorithm on sagemaker as its less prone to variance and seasonality but because of limited data points availability it led to subdued accuracy and hence adopted a better model than this.
    - Created model using fbprophet which is a third-party library from Facebook, for forecasting using Amazon SageMaker and went ahead with this approach as fbprophet was giving better accuracy and better results. Used Pytorch estimator in sagemaker for training and deployment of the model.
* Used spot instances as spot training can optimize the cost of training models up to 90% over on-demand instances.
* Deployed the ML model using Flask API.

.

**AWS Services used**

* Amazon SageMaker
* AWS API Gateway
* AWS S3 – Storing the Backup
* AWS Lambda
* AWS CloudWatch

**Solution Outcome**

* With all the customizations and hyperparameters tuning we selected a model with which we were able to forecast the sales with up to 93% accuracy. This helped to optimize manpower at sales channels. This helped the production team in adjusting the production plan as per the forecast.
* After running various models and trying various algorithms we reached to the decision of selecting one i.e., fbprophet.
* Reduced model training and deployment costs by using spot instances

**Architecture Diagram**

A picture containing chart

Description automatically generated

**How AWS services helped in building the model for sales Forecasting**

**Amazon SageMaker**

* Amazon SageMaker is used to create and manage Jupyter notebooks that were used to prepare and process data and to train and deploy the machine learning models.
* Amazon SageMaker High Power GPU Instance used for training of fbprophet Model
* Model is optimized using Amazon SageMaker Hyper Parameter Optimization

**AWS Lambda to handle the backend API calls**

It helped to initialize and validate the input and acted as the backend of the whole task. AWS Lambda lets us run code without provisioning or managing servers. Also, it helped to connect with various AWS API’s to acquire various insights from the inputs.

**Amazon API Gateway**

Amazon API Gateway is an AWS service for creating, publishing, maintaining, monitoring, and securing REST, HTTP, and WebSocket APIs at any scale.

**Amazon S3** **to store CSV raw documents**

It is an object storage service that offers industry-leading scalability, data availability, security, and performance. In this solution it.

**Amazon CloudWatch**

Amazon CloudWatch monitors your Amazon Web Services (AWS) resources and the applications you run on AWS in real time. You can use CloudWatch to collect and track metrics, which are variables you can measure for your resources and applications.