

**Site Profitability Analysis**

**Version: v3.3**

**August 2020**

**

|  |
| --- |
| **IMPORTANT NOTICE**  This document constitutes confidential and proprietary information of Subex Limited. The contents of this document may be accessed and/or used solely by a licensee of Subex Limited software product(s) and solely in connection with the licensee’s authorized use of such product(s), or as otherwise expressly permitted by Subex Limited in writing. All other uses are prohibited. This document may not in any event be disclosed to any third party without the prior written authorization of Subex Limited. |
| Copyright © 2006-2019 Subex Limited.  All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or information storage and retrieval systems for any purpose without the express written permission of Subex Limited. |

|  |  |
| --- | --- |
| **Project** | **Data Analytics: MTN Iran** |
| Document Author(s) | Anagha |
| Approved by | Anil Dilipkumar Wadhwani |
| Distribution List | MTNI, SAEI, Subex |
| Date | 02-Mar-2020 |

**Revision History**

| **Ver.** | **Date** | **Modified**  **By** | **Pages**  **Affected** | **Remarks** |
| --- | --- | --- | --- | --- |
| 1.0 | 02-Mar-2020 | Anagha P | All | Initial Draft |
| 2.4 | 26-July-2020 | Payam Mahmoudi | Business requirements/Data Flow | Adding new requirements/updating ERM and cost logic/Adding Dashboard traceability |
| 3.0 | 1-Aug-2020 | Negin Fazel | Data Flow Diagram | Updating data flow |
| 3.1 | 12-Aug-2020 | Payam Mahmoudi | Sample Output | Updating Screenshot of the dashboard/Updating aggregation tables/Updating data flow/Updating columns for input tables |
| 3.2 | 15-Aug-2020 | Payam Mahmoudi | Business Requirements | Adding new requirements/Adding columns for feasibility and data availability |
| 3.3 | 22-Aug-2020 | Payam Mahmoudi | Data Flow Diagram | Updating output table details/input table details |
| 3.4 | 24-Aug-2020 | Nasim Sharafi | Data Flow Diagram | Updating input table details |

**Acronyms**

|  |  |
| --- | --- |
| **Acronym** | **Expansion** |
| SL | Subex Limited |
| EDL | Enterprise Data Lake |

Table of Contents

Contents

[1. Scope of the Analysis 1-5](#_Toc40629087)

[2. Objective 1-6](#_Toc40629088)

[3. Business Requirements 1-7](#_Toc40629089)

[4. Business Benefit 1-9](#_Toc40629090)

[5. Technical Solution 1-10](#_Toc40629091)

[5.1 Data Collection 1-11](#_Toc40629092)

[5.2 Data Flow Diagram 1-11](#_Toc40629093)

[5.3 Tools and Techniques 1-22](#_Toc40629094)

[5.4 Data Processing (Exploratory Data Analysis) 1-23](#_Toc40629095)

[5.5 Modeling 1-23](#_Toc40629096)

[6. Sample Output 1-25](#_Toc40629100)

# Scope of the Analysis

This section captures the scope for analysis of the given use-case.

|  |  |
| --- | --- |
| **Name** | Site Profitability |
| **Stakeholders** | Marketing, S&D, Finance |
| **In – Scope** | Business requirements as captured in this document (as mentioned in section 3) |
| **Out – Scope** | Business requirements not part of this document (as mentioned in section 3) |
| **Assumptions & Pre-requisites** | At Present, summary datasets from BIB have been used for the analysis. All the relevant data sources and KPIs to be part of EDL |

# Objective

The objective of site profitability analysis is to provide detailed visibility at site level with respect to overall cost, revenue, and profit so that business teams can take calculated data driven business decisions to optimize capex and opex allocation for network sites.

# 3.Business Requirements

This section captures all the business requirements as raised by business teams for detailed Bolton analysis.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sr. No.** | **Business Requirement** | **Status** | **Remarks** | **Dashboard Traceability** | **Feasibility** | **Data Availability** |
| 1 | Identify profitable and non-profitable sites | Completed | Dashboard has been updated | Available in “Profitable & Non Profitable Sites” dashboard | Feasible | Available |
| 2 | Add regional level analysis | Completed | Dashboard has been updated | Available in “Geospatial View” dashboard | Feasible | Available |
| 3 | Identify impact of different cost components | Completed | Dashboard has been updated | Available in “Profitability Analysis” dashboard | Feasible | Available |
| 4 | Measure revenue at service level for profitable and non-profitable sites | Completed | Dashboard has been updated | Available in “Profitable & Non Profitable Sites” dashboard | Feasible | Available |
| 5 | Having whole details when using a specify filter | Completed | Dashboard has been updated | Available in all dashboards | Feasible | Available |
| 6 | For a non-profitable which of data/voice/sms has more effect | Completed | Dashboard has been updated | Available in “Profitable & Non Profitable Sites” dashboard | Feasible | Available |
| 7 | Exporting data in CSV format for revenue per site | Completed | Can be exported in tableau |  | Feasible | Available |
| 8 | Add information about planned sites | Completed | Dashboard has been updated | Status of the site is available as a filter (can see information just for planned sites) | Feasible | Available |
| 9 | Add GIS site type information | Completed | Dashboard has been updated | GIS site type is available as a filter | Feasible | Available |
| 10 | Adding profitable & non profitable percentage based on above( GIS and radio type filters) | Completed | Dashboard has been updated | Available in “Profitability Analysis” dashboard | Feasible | Available |
| 11 | Add Radio site type information | Completed | Dashboard has been updated | Radio site type is available as a filter | Feasible | Available |
| 12 | Add site status information | Completed | Dashboard has been updated | Site status is available as a filter | Feasible | Available |
| 13 | Add SIM technology usage(3G/4G) | Completed | Dashboard has been updated | Handset type is available as a filter | Feasible | Available |
| 14 | Excluding down  sites from nonprofitable | Completed | Dashboard has been updated | A filter is available to exclude down sites | Feasible | Available |
| 15 | Add percentage of high value customer of each site | Completed | Dashboard has been updated | Available in “Profitable & Non Profitable Sites” dashboard as a column | Feasible | Available |
| 16 | SIM Registration with type of SIM (prepaid or postpaid) | Completed | Dashboard has been updated | Available in “Profitable & Non Profitable Sites” dashboard as a column | Feasible | Available |
| 17 | Revenue and cost should be change based on selecting province, city or other filters | Completed | Dashboard has been updated | Available in all dashboards | Feasible | Available |
| 18 | BCG matrix with filter on radio and GIS site type | Completed | Dashboard has been updated | Available in “Geospatial View” dashboard | Feasible | Available |
| 19 | High value customer% of each site | Completed | Dashboard has been updated | Available in “Profitable & Non Profitable Sites” dashboard | Feasible | Available |
| 20 | Put a comment in the table to describe ERM value is dynamic value(ERM is different for each month) | Completed | Input table details have been updated |  | Feasible | Available |
| 21 | Add anomaly detection base on group of sites | In Progress |  |  | Feasible | Available |
| 22 | Add Alerting system to find abnormal sites in a site segment group | In Progress |  |  | Feasible | Available |
| 23 | Site whitelisting feature to track a group of selected sites | Pending |  |  | Not Feasible | Not Available |
| 24 | Calculate dormant users per each sites | Pending |  |  | Not Feasible | Not Available |
| 25 | Add 10 selected provinces in 2020 and compare their main KPI and also their score in each month | Pending |  |  | Feasible | Not Available |
| 26 | Ability to track  new sites 2 months after being On air | Pending |  |  | Feasible | Not Available |
| 27 | Getting a list of new sites and remove them from non-profitable | Pending |  |  | Feasible | Not Available |
| 28 | Network analysis based on hourly levels | Pending |  |  | Not Feasible | Not Available |
| 29 | Profit vs. cost graph in bottom right: doesn’t mean anything and can be removed | Pending | All stakeholders should agree |  | Feasible | Available |
| 30 | Can we zoom and see city and lack level in the profitability map | Pending | Currently map is offline |  | Not Feasible | Available |
| 31 | Add change percentage of each site VS previous months based on their margin | Pending |  |  | Feasible | Not Available |
| 32 | Add main KPI charts trend and exact figures (data, voice,…) | Pending | Need clarification about exact KPIs |  | Feasible | Not Available |
| 33 | Add NWK KPI(payload, throuput,VLR,ERLANG,…) | Pending |  |  | Feasible | Not Available |
| 34 | Add number of actual and target site for new site,modernization,down site for each province in 2020 | Pending |  |  | Feasible | Not Available |
| 35 | Tracking relocations-before & after | Pending |  |  | Not Feasible | Not Available |
| 36 | Possibility to select the value of margin for example show the sites with margin more than x | Pending |  |  | Feasible | Available |
| 37 | Attached data subs per site, throuput,payload,HVC | Pending |  |  | Feasible | Not Available |
| 38 | We need to see voice dormant users / data dormant users of the site | Pending |  |  | Feasible | Not Available |
| 39 | We need  list of all negative sites with margin more than x | Pending |  |  | Feasible | Available |
| 40 | Adding the population of the region that site is located if possible and have enough accuracy  to find potential capacity of new subs | Pending |  |  | Not Feasible | Not Available |
| 41 | Adding DATA and VOICE capacity | Pending |  |  | Not Feasible | Not Available |
| 42 | Adding device registration with type of device (TDD or FDD) | Pending |  |  | Feasible | Not Available |
| 43 | Recharge and echarge per merchant | Pending |  |  | Not Feasible | Not Available |
| 44 | Removing IMEI Count column | Pending |  |  | Feasible | Available |

# 4.Business Benefit

This section lists down the various business benefits that the business teams can look for, with respect to this use-case…

1. Identify sites which are more profitable or have high value customers and take necessary steps to further Improve customer experience and network availability.
2. Allocate capex and opex costs more effectively since detailed KPIs are available at site level.
3. Design network strategy not only as per network KPIs but also considering the cost, revenue and profits involved.
4. Associate revenue of customer across the cell sites. Classify the cell sites based on the data, voice and SMS usage behaviour.

# 5.Technical Solution

This section covers detailed methodology followed for this use-case, as captured in below mentioned sub-sections

1. Data Collection
2. Data Flow Diagram

* EDL Table details
* Input table details
* Output table details

1. Tools & Techniques
2. Data Pre-processing (EDA)
3. Modelling

## 5.1 Data Collection

Data collection has been done based on the requirements raised by business teams as captured in section 3 of this document. Following are the data sources that have been identified for the use-case.

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **SOURCE Table name** | **Actual Source Name** | **EDL Table name** |
| Site Information | CELL CFG | FACT/MAPS/CCN/FIN | dim\_base\_station,  dl\_msisdn\_cellsite\_voice\_gprs\_sms |
| CCN VOICE CDR,CCN SMS CDR,CCN GPRS CDR |  |
| Excel Sheet : Site Cost | mt\_site\_cost\_finance |
| Site Details | Excel Sheet : Site Report | dim\_base\_station\_details |
| Sim information | ABILITY GSM SIMS MASTER | CBS | edl\_simtype |
| Handset information | GEMALTO OTFA DEVICE INFORMATION | GEMALTO | edl\_src\_gemalto\_ota\_device |
| Customer Segmentation | CCN SDP DUMP MAIN  AILITY CB PACKAGE  ABILITY GSM SERVICE MAST  ABILITY CB CHANGE OWNER HISTORY | CBS/CCN | edl\_fct\_subs\_cvs\_snapm\_ir |
| BIB.DIM\_CUSTOMER\_SEGMENT | BIB |

## 5.2 Data Flow Diagram

This section provides an overview of data flow for the use-case along with details of EDL Tables, Input Tables and Output Tables. Currently the input tables are from BIB, once the actual data sources are integrated, data flow diagram shall be updated.



**EDL TABLE DETAILS:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SL NO.** | **DESCRIPTION** | **TABLE TYPE** | **EDL TABLE NAME** | **LOGIC** |
| 1 | SITEID LEVEL COST INFORMATION | INPUT | mt\_site\_cost\_finance | This will be direct input from business team |
| 2 | VOICE-GPRS-SMS USAGE TABLE | INPUT | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Available directly |
| 3 | CUSTOMER SEGMENT AND CONTRACT TYPE | INPUT | edl\_fct\_subs\_cvs\_snapm\_ir | Available directly |
| 4 | ERM TABLE | INPUT | edl\_erm (hardcode) | This will be direct input from business team |
| 5 | CELL SITE INFORMATION | INPUT | dim\_base\_station  dim\_base\_station\_details | Available directly |
| 6 | DEVICE INFORMATION | INPUT | edl\_src\_gemalto\_ota\_device | Available directly |
| 7 | DAILY/WEEKLY/MONTHLY SITEID LEVEL COST INFORMATION | AGGREGATED | dl\_site\_cost\_finance\_enriched,  mt\_site\_cost\_finance\_enriched,  wk\_site\_cost\_finance\_enriched | SITEID level aggregated data. |
| 8 | AGGREGATED USAGE TABLE | AGGREGATED | dl\_site\_usage\_enriched  wk\_site\_usage\_enriched  mt\_site\_usage\_enriched | SITEID level aggregated data. |
| 9 | VOICE/DATA/SMS REVENUE TABLE | AGGREGATED | dl\_site\_usage\_enriched  wk\_site\_usage\_enriched  mt\_site\_usage\_enriched | Data Revenue = Data Usage \* Data ERM  Voice Revenue= Voice usage \* Voice ERM  SMS Revenue = SMS usage \* SMS ERM  **Total Revenue = Data Revenue + Voice Revenue + SMS Revenue** |
| 10 | SITE PROFITABILITY MASTER TABLE | OUTPUT | cellsite\_profitability\_dl\_wk\_mt\_combined  cellsite\_profitability\_dl\_anomaly | Profit % = (Total Revenue – Total Cost) / Total Revenue |

**INPUT TABLE DETAILS:**

**SITEID LEVEL COST INFORMATION:**

|  |  |  |  |
| --- | --- | --- | --- |
| **COLUMN NAME** | **DATA TYPE** | **CONFIDENTIALITY** | **DATA QUALITY CHECK** |
| siteid | String | Confidential | No null values, No missing values |
| lease\_cost | Double | Confidential | No null values, No missing values |
| Fuel | Double | Confidential | No null values, No missing values |
| guards | Double | Confidential | No null values, No missing values |
| ms\_variable | Double | Confidential | No null values, No missing values |
| ms\_fixed | Double | Confidential | No null values, No missing values |
| electricity | Double | Confidential | No null values, No missing values |
| tx\_l2 | Double | Confidential | No null values, No missing values |
| internet\_cost | Double | Confidential | No null values, No missing values |
| vpls\_cost | Double | Confidential | No null values, No missing values |
| space\_and\_power | Double | Confidential | No null values, No missing values |
| total\_transmission\_cost | Double | Confidential | No null values, No missing values |

**VOICE-GPRS-SMS USAGE TABLE:**

|  |  |  |  |
| --- | --- | --- | --- |
| **COLUMN NAME** | **DATA TYPE** | **CONFIDENTIALITY** | **DATA QUALITY CHECK** |
| date\_key | Double | Confidential | No null values, No missing values |
| msisdn | Double | Confidential | No null values, No missing values |
| base\_station | Double | Confidential | No null values, No missing values |
| voice\_duration | Double | Confidential | No null values, No missing values |
| voice\_cnt | Double | Confidential | No null values, No missing values |
| datasession\_duration | Double | Confidential | No null values, No missing values |
| data\_cnt | Double | Confidential | No null values, No missing values |
| data\_bytes\_qty | Double | Confidential | No null values, No missing values |
| sms\_cnt | Double | Confidential | No null values, No missing values |



**ERM TABLE:**

|  |  |  |  |
| --- | --- | --- | --- |
| **COLUMN NAME** | **DATA TYPE** | **CONFIDENTIALITY** | **DATA QUALITY CHECK** |
| date\_key | Date | Confidential | No null values, No missing values |
| data\_erm | Double | Confidential | No null values, No missing values |
| voice\_erm | Double | Confidential | No null values, No missing values |
| sms\_erm | Double | Confidential | No null values, No missing values |

**CELL SITE INFORMATION TABLE:**

**dim\_base\_station**

|  |  |  |  |
| --- | --- | --- | --- |
| **COLUMN NAME** | **DATA TYPE** | **CONFIDENTIALITY** | **DATA QUALITY CHECK** |
| base\_station\_key | Double | Confidential | No null values, No missing values |
| base\_station\_cd | String | Confidential | No null values, No missing values |
| siteid | String | Confidential | No null values, No missing values |
| base\_station\_lat\_txt | String | Confidential | No null values, No missing values |
| base\_station\_long\_txt | String | Confidential | No null values, No missing values |
| base\_station\_site\_nm | String | Confidential | No null values, No missing values |
| base\_station\_city\_nm | String | Confidential | No null values, No missing values |
| base\_station\_province\_nm | String | Confidential | No null values, No missing values |
| base\_station\_region\_nm | String | Confidential | No null values, No missing values |
| base\_station\_technology\_txt | String | Confidential | No null values, No missing values |

**CELL SITE INFORMATION TABLE:**

**dim\_base\_station\_details**

|  |  |  |  |
| --- | --- | --- | --- |
| **COLUMN NAME** | **DATA TYPE** | **CONFIDENTIALITY** | **DATA QUALITY CHECK** |
| site\_id | String | Confidential | No null values, No missing values |
| latitude | String | Confidential | No null values, No missing values |
| longitude | String | Confidential | No null values, No missing values |
| city | String | Confidential | No null values, No missing values |
| province | String | Confidential | No null values, No missing values |
| gis\_site\_type | String | Confidential | No null values, No missing values |
| township | String | Confidential | No null values, No missing values |
| radio\_site\_type | String | Confidential | No null values, No missing values |
| region | String | Confidential | No null values, No missing values |
| physical\_site\_status | String | Confidential | No null values, No missing values |
| 2g\_status | String | Confidential | No null values, No missing values |
| 3g\_service\_status | String | Confidential | No null values, No missing values |
| lte\_fdd\_service\_status | String | Confidential | No null values, No missing values |
| bts\_type\_current | String | Confidential | No null values, No missing values |
| support\_height | String | Confidential | No null values, No missing values |
| support\_type | String | Confidential | No null values, No missing values |
| physical\_on\_air\_date | String | Confidential | No null values, No missing values |

**DEVICE INFORMATION:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Data Type** | **CONFIDENTIALITY** | **Data Quality Check** |
| msisdn | | string | Confidential | No null values, No missing values |
| imei | string | Confidential | No null values, No missing values |
| imsi | double | Confidential | No null values, No missing values |
| last\_detection | string | Confidential | No null values, No missing values |
| brand\_name | string | Confidential | No null values, No missing values |
| model\_name | string | Confidential | No null values, No missing values |
| device\_category | string | Confidential | No null values, No missing values |
| os\_name | string | Confidential | No null values, No missing values |
| two\_g | String | Confidential | No null values, No missing values |
| three\_g | String | Confidential | No null values, No missing values |
| three\_gp | String | Confidential | No null values, No missing values |
| lte | String | Confidential | No null values, No missing values |
| msisdn\_nsk | double | Confidential | No null values, No missing values |

CUSTOMER SEGMENT AND CONTRACT TYPE:

|  |  |  |  |
| --- | --- | --- | --- |
| **COLUMN NAME** | **DATA TYPE** | **CONFIDENTIALITY** | **DATA QUALITY CHECK** |
| customer\_segment\_txt | String | Confidential | No null values, No missing values |
| payment\_option\_cd | String | Confidential | No null values, No missing values |
| date\_key | Date | Confidential | No null values, No missing values |
| customer\_status | String | Confidential | No null values, No missing values |
| msisdn\_nsk | Bigint | Confidential | No null values, No missing values |
| account\_link\_code\_n | Double | Confidential | No null values, No missing values |
| subs\_status\_code | String | Confidential | No null values, No missing values |
| account\_activated\_ind | String | Confidential | No null values, No missing values |
| registration\_dt | String | Confidential | No null values, No missing values |
| last\_cow\_date | String | Confidential | No null values, No missing values |
| actual\_activation\_dt | String | Confidential | No null values, No missing values |
| tenure\_months\_cnt | Int | Confidential | No null values, No missing values |
| tenure\_days\_cnt | Int | Confidential | No null values, No missing values |
| tenure\_rank\_nr | Double | Confidential | No null values, No missing values |
| tenure\_customer\_segment\_cd | String | Confidential | No null values, No missing values |
| all\_num\_subs | Double | Confidential | No null values, No missing values |
| rgs\_3months\_amt | Double | Confidential | No null values, No missing values |
| arpu\_amt | Double | Confidential | No null values, No missing values |
| customer\_segment\_key | Double | Confidential | No null values, No missing values |
| cvs\_customer\_segment\_key | Double | Confidential | No null values, No missing values |
| arpu\_rank\_nr | Double | Confidential | No null values, No missing values |
| customer\_segment\_cd | String | Confidential | No null values, No missing values |
| cvs\_score\_value\_qty | Double | Confidential | No null values, No missing values |
| cvs\_rank\_nr | Double | Confidential | No null values, No missing values |
| cvs\_customer\_segment\_cd | String | Confidential | No null values, No missing values |
| dw\_subpart | Double | Confidential | No null values, No missing values |
| batch\_id | Double | Confidential | No null values, No missing values |
| create\_dt | String | Confidential | No null values, No missing values |

**OUTPUT TABLES DETAILS:**

**cellsite\_profitability\_dl\_wk\_mt\_combined**

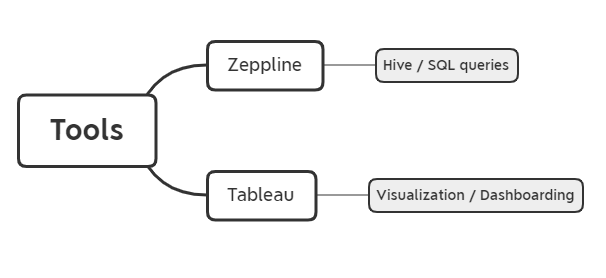
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **COLUMN NAME** | **EDL TABLE NAME** | **KPI TYPE** | **LOGIC** | **KPI DESCRIPTION** |
| site\_id | mt\_site\_cost\_finance | Basic | - | Unique ID for the site |
| date\_key1 | mt\_site\_cost\_finance | Derived | weekofyear(date) | Week number of the year. |
| date\_key | mt\_site\_cost\_finance | Basic | - | Date |
| msisdn\_cnt | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | count(distinct msisdn) | Number of people latched to the particular site. |
| voice\_duration\_mnt | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | - | Total voice duration in Minutes. |
| voice\_cnt | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Basic | - | count of voice calls |
| datasession\_duration | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Basic | - | Total duration of the data sessions |
| data\_cnt | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Basic | - | count of data sessions |
| data\_usage\_mb | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | - | Total data usage in MB |
| sms\_cnt | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Basic | - | Count of SMS |
| base\_station\_region\_name | dim\_base\_station | Basic | - | Region name of the site |
| base\_station\_province\_name | dim\_base\_station | Basic | - | Province name of the site |
| base\_station\_city\_name | dim\_base\_station | Basic | - | City name of the site |
| base\_station\_lat\_txt | dim\_base\_station | Basic | - | Latitude of the site |
| base\_station\_long\_txt | dim\_base\_station | Basic | - | Longitude of the site |
| site\_technology | dim\_base\_station | Derived | max(if((base\_station\_technology\_txt='LTE') or (base\_station\_technology\_txt='TDD-LTE') or (base\_station\_technology\_txt='WIMAX'),1,0)) as t\_4G, max(if(base\_station\_technology\_txt='3G',1,0)) as t\_3G, max(if(((base\_station\_technology\_txt='2G') or (base\_station\_technology\_txt='GSM')),1,0)) as t\_2G, max(if((base\_station\_technology\_txt='NA')or (base\_station\_technology\_txt='') or (base\_station\_technology\_txt is null),1,0)) as no\_info  then using this columns  if((t\_4G=1 and t\_3G=0 and t\_2G=0 and no\_info=0),'4G\_only', if((t\_4G=1 and t\_3G=1 and t\_2G=0 and no\_info=0 ),'4G\_3G', if((t\_4G=1 and t\_3G=0 and t\_2G=1 and no\_info=0 ),'4G\_2G', if((t\_4G=1 and t\_3G=1 and t\_2G=1 and no\_info=0 ),'4G\_3G\_2G', if((t\_4G=0 and t\_3G=1 and t\_2G=0 and no\_info=0 ),'3G\_only', if((t\_4G=0 and t\_3G=1 and t\_2G=1 and no\_info=0 ),'3G\_2G', if((t\_4G=0 and t\_3G=0 and t\_2G=1 and no\_info=0 ),'2G\_only','others' ))))))) | Latest technology supported by the site |
| gis\_site\_type | dim\_base\_station\_details | Basic |  | Gis Site Type |
|  |  |  |  |  |
| township | dim\_base\_station\_details | Basic | - | Township of the site |
| radio\_site\_type | dim\_base\_station\_details | Basic | - | Radio site type |
| region | dim\_base\_station\_details | Basic | - | Region of the site |
| Physical\_site\_status | dim\_base\_station\_details | Basic | - | Status of the site |
| 2g\_status | dim\_base\_station\_details | Basic | - | 2G status of the site |
| 3g\_service\_status | dim\_base\_station\_details | Basic | - | 3G status of the site |
| lte\_fdd\_service\_status | dim\_base\_station\_details | Basic | - | LTE/FDD status of the site |
| bts\_type\_current | dim\_base\_station\_details | Basic | - |  |
| support\_height | dim\_base\_station\_details | Basic | - | Site’s height support |
| support\_type | dim\_base\_station\_details | Basic | - | Site’s support type |
| physical\_on\_air\_date | dim\_base\_station\_details | Basic | - | Site’s on air date |
| voice\_rev | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | voice\_duration\_mnt\*ERM | Voice revenue (ERM= 0.027 per minute and can be different in different months) |
| data\_rev | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | (data\_usage\_mb\*1024)\*ERM | Data Revenue ( For data ERM=578 per KB and can be different in different months) |
| sms\_rev | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | sms\_cnt\*ERM | SMS Revenue (For SMS ERM=129 and can be different in different months) |
| tot\_revenue | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | voice\_rev+ data\_rev + sms\_rev | Total revenue for usage. |
| lease\_cost | mt\_site\_cost\_finance | Basic | - | Lease cost |
| fuel | mt\_site\_cost\_finance | Basic | - | Fuel cost |
| guards | mt\_site\_cost\_finance | Basic | - | Guards cost |
| ms\_variable | mt\_site\_cost\_finance | Basic | - |  |
| ms\_fixed | mt\_site\_cost\_finance | Basic | - |  |
| electricity | mt\_site\_cost\_finance | Basic | - | Electricity cost |
| tx\_l2 | mt\_site\_cost\_finance | Basic | - |  |
| vpls\_cost | mt\_site\_cost\_finance | Basic | - | Vpls cost |
| space\_and\_power | mt\_site\_cost\_finance | Basic | - |  |
| total\_transmission\_cost | mt\_site\_cost\_finance | Basic | - | Total transmission cost |
| total\_sitecost | mt\_site\_cost\_finance | Derived | lease\_cost +fuel + guards+ms\_variable+ ms\_fixed + electricity+ total\_transmission\_cost | Total cost spend for the site |
| 4g\_handset | edl\_src\_gemalto\_ota\_device | Derived | count(distinct case when handset\_type='4G' then msisdn end) | Number of customers who their handset type is 4G |
| 3g\_handset | edl\_src\_gemalto\_ota\_device | Derived | count(distinct case when handset\_type='3G' then msisdn end) | Number of customers who their handset type is 3G |
| 2g\_handset | edl\_src\_gemalto\_ota\_device | Derived | count(distinct case when handset\_type='2G' then msisdn end) | Number of customers who their handset type is 2G |
| other\_handset | edl\_src\_gemalto\_ota\_device | Derived | count(distinct case when handset\_type='Others' then msisdn end) | Number of customers who their handset type is other |
| prepaid\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(distinct CASE WHEN payment\_option\_cd='P' THEN msisdn end) | Number of prepaid customers |
| postpaid\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(distinct CASE WHEN payment\_option\_cd='N' THEN msisdn end) | Number of postpaid customers |
| unknown\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(distinct CASE WHEN payment\_option\_cd='NA' THEN msisdn end) | Number of unknown contract type customers |
| top\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'TOP' then msisdn end) | Number of top customers |
| veryhigh\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'VERY HIGH' then msisdn end) | Number of very high customers |
| high\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'HIGH' then msisdn end) | Number of high customers |
| medium\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'MEDIUM' then msisdn end) | Number of medium customers |
| low\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'LOW' then msisdn end) | Number of low customers |
| newcomers\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'NEWCOMER' then msisdn end) | Number of newcommer customers |
| others\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt not in ('TOP','VERY HIGH','HIGH','MEDIUM','LOW','NEWCOMER') then msisdn end) | Number of other segment customers |
| voice\_usage\_bin | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | - | Voice usage is classified into different buckets (zero,low,medium,high,very high) |
| data\_usage\_bin | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | - | Data usage is classified into different buckets (zero,low,medium,high,very high) |
| sms\_usage\_bin | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | - | SMS usage is classified into different buckets (zero,low,medium,high,very high) |
| period\_type |  | Derived | Daily profiler tagged as- Daily, weekly profiler tagged as - weekly, Monthly profiler tagged as -Monthly | Indicator for the tag |

**cellsite\_profitability\_dl\_anomaly**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **COLUMN NAME** | **EDL TABLE NAME** | **KPI TYPE** | **LOGIC** | **KPI DESCRIPTION** |
| site\_id | mt\_site\_cost\_finance | Basic | - | Unique ID for the site |
| date\_key1 | mt\_site\_cost\_finance | Derived | weekofyear(date) | Week number of the year. |
| date\_key | mt\_site\_cost\_finance | Basic | - | Date |
| msisdn\_cnt | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | count(distinct msisdn) | Number of people latched to the particular site. |
| voice\_duration\_mnt | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | - | Total voice duration in Minutes. |
| voice\_cnt | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Basic | - | count of voice calls |
| datasession\_duration | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Basic | - | Total duration of the data sessions |
| data\_cnt | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Basic | - | count of data sessions |
| data\_usage\_mb | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | - | Total data usage in MB |
| sms\_cnt | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Basic | - | Count of SMS |
| base\_station\_region\_name | dim\_base\_station | Basic | - | Region name of the site |
| base\_station\_province\_name | dim\_base\_station | Basic | - | Province name of the site |
| base\_station\_city\_name | dim\_base\_station | Basic | - | City name of the site |
| base\_station\_lat\_txt | dim\_base\_station | Basic | - | Latitude of the site |
| base\_station\_long\_txt | dim\_base\_station | Basic | - | Longitude of the site |
| site\_technology | dim\_base\_station | Derived | max(if((base\_station\_technology\_txt='LTE') or (base\_station\_technology\_txt='TDD-LTE') or (base\_station\_technology\_txt='WIMAX'),1,0)) as t\_4G, max(if(base\_station\_technology\_txt='3G',1,0)) as t\_3G, max(if(((base\_station\_technology\_txt='2G') or (base\_station\_technology\_txt='GSM')),1,0)) as t\_2G, max(if((base\_station\_technology\_txt='NA')or (base\_station\_technology\_txt='') or (base\_station\_technology\_txt is null),1,0)) as no\_info  then using this columns  if((t\_4G=1 and t\_3G=0 and t\_2G=0 and no\_info=0),'4G\_only', if((t\_4G=1 and t\_3G=1 and t\_2G=0 and no\_info=0 ),'4G\_3G', if((t\_4G=1 and t\_3G=0 and t\_2G=1 and no\_info=0 ),'4G\_2G', if((t\_4G=1 and t\_3G=1 and t\_2G=1 and no\_info=0 ),'4G\_3G\_2G', if((t\_4G=0 and t\_3G=1 and t\_2G=0 and no\_info=0 ),'3G\_only', if((t\_4G=0 and t\_3G=1 and t\_2G=1 and no\_info=0 ),'3G\_2G', if((t\_4G=0 and t\_3G=0 and t\_2G=1 and no\_info=0 ),'2G\_only','others' ))))))) | Latest technology supported by the site |
| gis\_site\_type | dim\_base\_station\_details | Basic |  | Gis Site Type |
|  |  |  |  |  |
| township | dim\_base\_station\_details | Basic | - | Township of the site |
| radio\_site\_type | dim\_base\_station\_details | Basic | - | Radio site type |
| region | dim\_base\_station\_details | Basic | - | Region of the site |
| Physical\_site\_status | dim\_base\_station\_details | Basic | - | Status of the site |
| 2g\_status | dim\_base\_station\_details | Basic | - | 2G status of the site |
| 3g\_service\_status | dim\_base\_station\_details | Basic | - | 3G status of the site |
| lte\_fdd\_service\_status | dim\_base\_station\_details | Basic | - | LTE/FDD status of the site |
| bts\_type\_current | dim\_base\_station\_details | Basic | - |  |
| support\_height | dim\_base\_station\_details | Basic | - | Site’s height support |
| support\_type | dim\_base\_station\_details | Basic | - | Site’s support type |
| physical\_on\_air\_date | dim\_base\_station\_details | Basic | - | Site’s on air date |
| voice\_rev | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | voice\_duration\_mnt\*ERM | Voice revenue (ERM= 0.027 per minute and can be different in different months) |
| data\_rev | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | (data\_usage\_mb\*1024)\*ERM | Data Revenue ( For data ERM=578 per KB and can be different in different months) |
| sms\_rev | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | sms\_cnt\*ERM | SMS Revenue (For SMS ERM=129 and can be different in different months) |
| tot\_revenue | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | voice\_rev+ data\_rev + sms\_rev | Total revenue for usage. |
| lease\_cost | mt\_site\_cost\_finance | Basic | - | Lease cost |
| fuel | mt\_site\_cost\_finance | Basic | - | Fuel cost |
| guards | mt\_site\_cost\_finance | Basic | - | Guards cost |
| ms\_variable | mt\_site\_cost\_finance | Basic | - |  |
| ms\_fixed | mt\_site\_cost\_finance | Basic | - |  |
| electricity | mt\_site\_cost\_finance | Basic | - | Electricity cost |
| tx\_l2 | mt\_site\_cost\_finance | Basic | - |  |
| vpls\_cost | mt\_site\_cost\_finance | Basic | - | Vpls cost |
| space\_and\_power | mt\_site\_cost\_finance | Basic | - |  |
| total\_transmission\_cost | mt\_site\_cost\_finance | Basic | - | Total transmission cost |
| total\_sitecost | mt\_site\_cost\_finance | Derived | lease\_cost +fuel + guards+ms\_variable+ ms\_fixed + electricity+ total\_transmission\_cost | Total cost spend for the site |
| 4g\_handset | edl\_src\_gemalto\_ota\_device | Derived | count(distinct case when handset\_type='4G' then msisdn end) | Number of customers who their handset type is 4G |
| 3g\_handset | edl\_src\_gemalto\_ota\_device | Derived | count(distinct case when handset\_type='3G' then msisdn end) | Number of customers who their handset type is 3G |
| 2g\_handset | edl\_src\_gemalto\_ota\_device | Derived | count(distinct case when handset\_type='2G' then msisdn end) | Number of customers who their handset type is 2G |
| other\_handset | edl\_src\_gemalto\_ota\_device | Derived | count(distinct case when handset\_type='Others' then msisdn end) | Number of customers who their handset type is other |
| prepaid\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(distinct CASE WHEN payment\_option\_cd='P' THEN msisdn end) | Number of prepaid customers |
| postpaid\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(distinct CASE WHEN payment\_option\_cd='N' THEN msisdn end) | Number of postpaid customers |
| unknown\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(distinct CASE WHEN payment\_option\_cd='NA' THEN msisdn end) | Number of unknown contract type customers |
| top\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'TOP' then msisdn end) | Number of top customers |
| veryhigh\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'VERY HIGH' then msisdn end) | Number of very high customers |
| high\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'HIGH' then msisdn end) | Number of high customers |
| medium\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'MEDIUM' then msisdn end) | Number of medium customers |
| low\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'LOW' then msisdn end) | Number of low customers |
| newcomers\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt = 'NEWCOMER' then msisdn end) | Number of newcommer customers |
| others\_subs | edl\_fct\_subs\_cvs\_snapm\_ir | Derived | count(DISTINCT case when customer\_segment\_txt not in ('TOP','VERY HIGH','HIGH','MEDIUM','LOW','NEWCOMER') then msisdn end) | Number of other segment customers |
| voice\_usage\_bin | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | - | Voice usage is classified into different buckets (zero,low,medium,high,very high) |
| data\_usage\_bin | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | - | Data usage is classified into different buckets (zero,low,medium,high,very high) |
| sms\_usage\_bin | dl\_msisdn\_cellsite\_voice\_gprs\_sms | Derived | - | SMS usage is classified into different buckets (zero,low,medium,high,very high) |
| current\_profit | dl\_msisdn\_cellsite\_voice\_gprs\_sms mt\_site\_cost\_finance | Derived | nvl(tot\_revenue,0)-nvl(total\_sitecost,0) | Current profit |
| previous\_profit | dl\_msisdn\_cellsite\_voice\_gprs\_sms mt\_site\_cost\_finance | Derived | lag(nvl(tot\_revenue,0)-nvl(total\_sitecost,0)) over (partition by site\_id,period\_type,dayofweek(date\_key) order by date\_key) | Median of previous same weekdays profits |
| period\_type |  | Derived | Daily profiler tagged as- Daily, weekly profiler tagged as - weekly, Monthly profiler tagged as -Monthly | Indicator for the tag |
| percent\_drop |  | Derived | if((lag(nvl(tot\_revenue,0)-nvl(total\_sitecost,0)) over (partition by site\_id,period\_type,dayofweek(date\_key) order by date\_key))<0,  (nvl(tot\_revenue,0)-nvl(total\_sitecost,0)-(lag(nvl(tot\_revenue,0)-nvl(total\_sitecost,0)) over (partition by site\_id,period\_type,dayofweek(date\_key) order by date\_key)))/((-1)\*(lag(nvl(tot\_revenue,0)-nvl(total\_sitecost,0)) over (partition by site\_id,period\_type,dayofweek(date\_key) order by date\_key))) ,  (nvl(tot\_revenue,0)-nvl(total\_sitecost,0)-(lag(nvl(tot\_revenue,0)-nvl(total\_sitecost,0)) over (partition by site\_id,period\_type,dayofweek(date\_key) order by date\_key)))/(lag(nvl(tot\_revenue,0)-nvl(total\_sitecost,0)) over (partition by site\_id,period\_type,dayofweek(date\_key) order by date\_key))) | Change in profit |

## 5.3 Tools and Techniques

This section provides an overview of tools used for the use-case



## 5.4 Data Processing (Exploratory Data Analysis)

Not Applicable for this use case

## Modeling

### 5.5.1 Algorithms

Not applicable for this use case.

### 5.5.2 Model Evaluation and Selection

Not applicable for this use case.

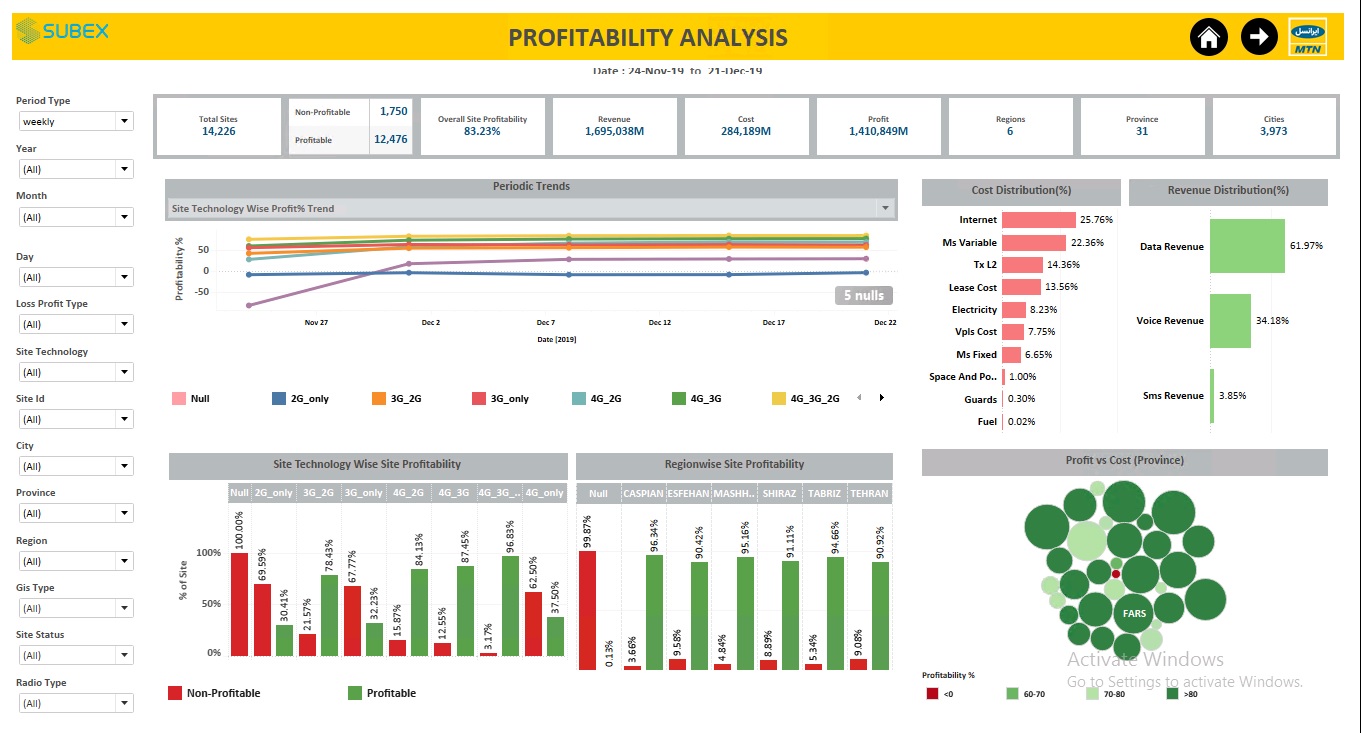
### 5.5.3 Model monitoring and governance framework

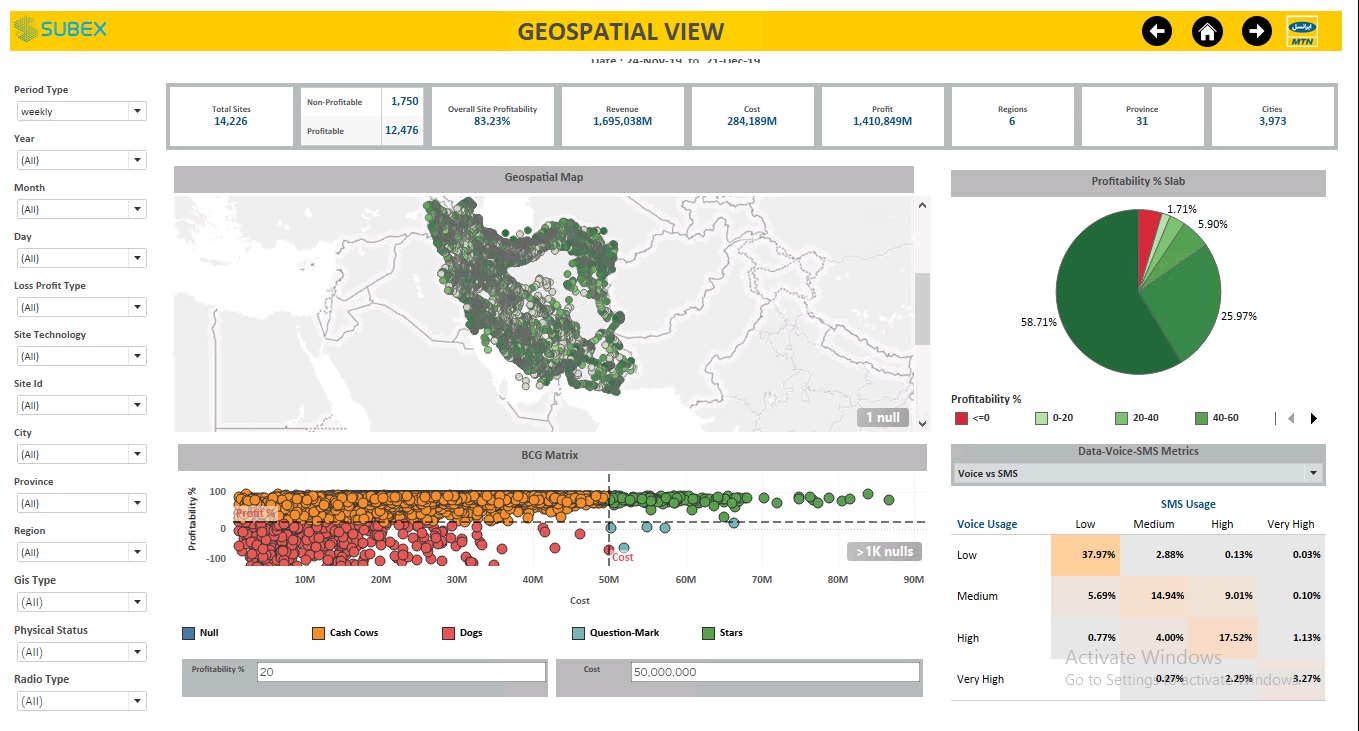
#### Refresh Rate

Dashboards will be refreshed every day. Predicted Resolution, Calls Count and Agent scoring are evaluated against the newly available data and the corresponding statistics are captured in the dashboard.

# 6.Sample Output

This section provides the details of the sample output for the use-case





**BCG Matrix Quadrants:**

#### *Question Marks*

#### Question marks are low-share business units in high-growth markets. They require cash to hold their share, let alone increase it. The company needs to think hard about question marks – which ones should be built into stars, and which ones should be phased out? Question marks have the following characteristics:

* *Low relative market share in a relatively young but promising market (growing)*
* *Potential of becoming stars if the market share can be increased*
* *If necessary market share is not reached, question marks are likely to turn into dogs as soon as the market gets more mature*

#### *Stars*

Stars are high-growth, high-share businesses or products. They often need heavy investment to finance their rapid growth. Eventually, their growth will slow down, and they will turn into cash cows. Stars have the following characteristics:

* *High market share in a promising market*
* *To turn a star into a future cash cow, heavy investment is needed to fight competition and expand market share*

#### *Cash Cows*

Cash cows are low-growth, high-share businesses or products. These established and successful SBUs need less investment to maintain their market share. As a result, they produce cash that the company uses to pay its bills and to support other SBUs that need investment. As we have learned, question marks and stars require heavy investment, which usually comes from the profitable cash cows. Cash cows have the following characteristics:

* *High market share in a slowly growing or mature market*
* *Create the highest cash flow*
* *No further investment should be undertaken due to limited or non-existent growth potential*
* *The company should try to “milk” the cash cows as long as possible.*

#### *(Poor) Dogs*

Dogs are low-growth, low-share businesses and products. They may generate enough cash to maintain themselves, but do not promise to be large sources of cash flow. Dogs have the following characteristics:

* *Low relative market share in a slowly growing or declining market*
* *Products do mostly not generate large profit and may usually just break even*
* *The company should divest dogs, as these products have a negative effect on the overall profitability of the company. Instead of carrying dogs along, the company should better focus on products or SBUs with greater potential.*

**Data Voice SMS Matrix**

These values are calculated based on the percentile

***Data Usage Bucket***

Low: data usage(mb) <= 10 percentile

Medium: data usage<=60 percentile and data usage>10 percentile

High: data usage <= 95 percentile and data usage > 60 percentile

Very High: > 95 percentile

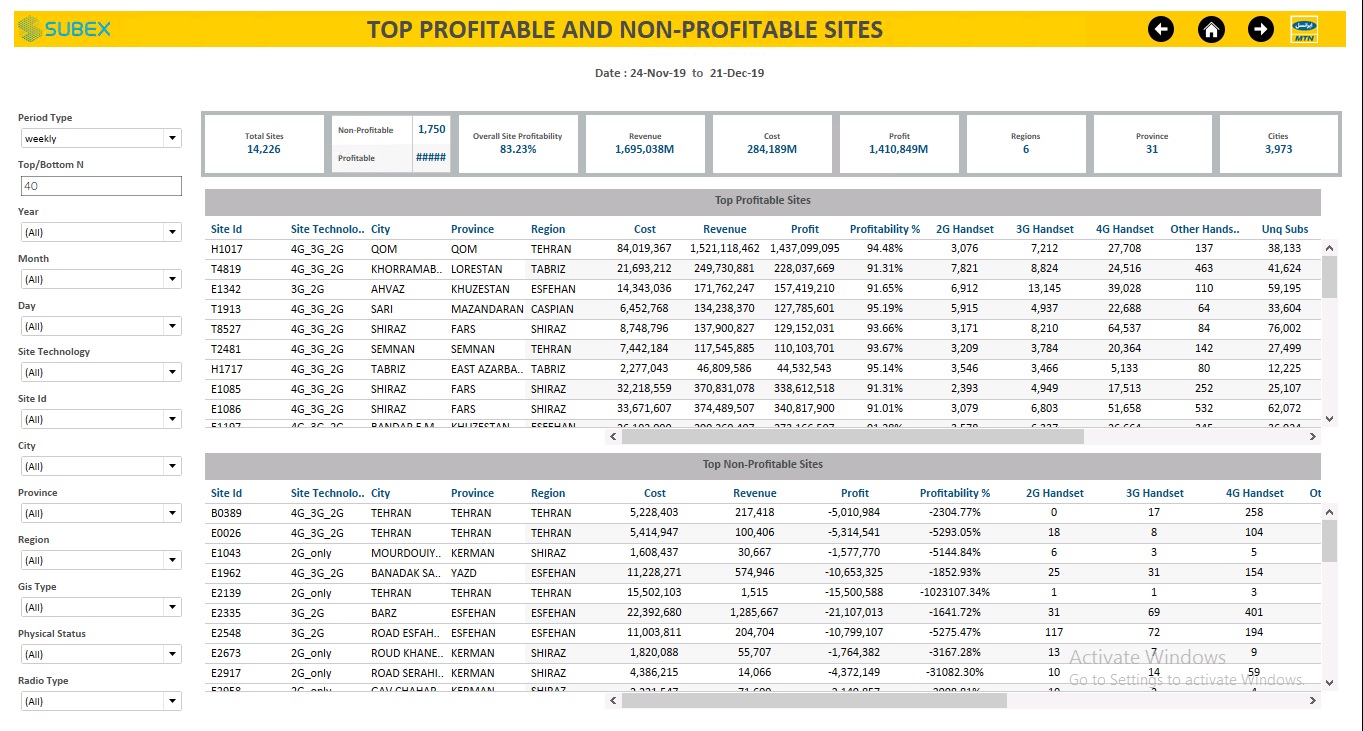
***Voice Usage Bucket***

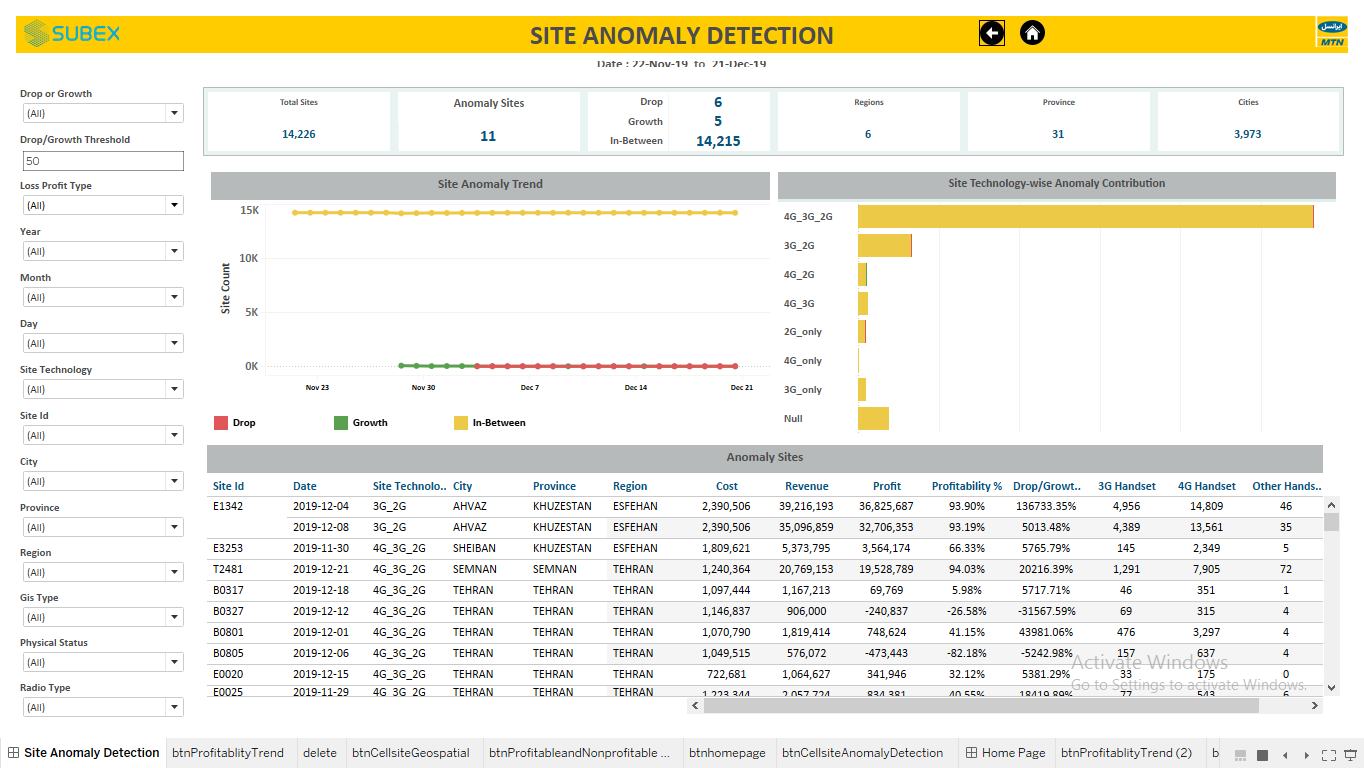
Low: voice usage(min) <= 50 percentile

Medium: voice usage<=75 percentile and data usage>50 percentile

High: voice usage <= 90 percentile and data usage > 75 percentile

Very High: >90 percentile





**Anomaly Calculation:** For a site, if anomaly occurs on 4th Monday of the month then we will take Median of profitability of the previous 3 Mondays and subtract it with the profitability of the current month Monday. If the final value comes out to be positive, then it is growth or else it is drop if the value is negative.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Week** | **Date** | **Profit** | **Median of Profit (last 3 Mondays)** | **Growth/Drop %** |
| 1 | Monday | 2000 | 0 |  |
| 2 | Monday | 1000 | 0 |  |
| 3 | Monday | 3500 | 0 |  |
| 4 | Monday | 2730 | 2000 (Median of 2000,1000,3500) | 37% (Profit-Median Profit)/Median Profit) |