**Derive and visualize Key Performance Indicators (KPIs) from BMTC’s ticketing data**

**Introduction to BMTC Ticketing Data:**

BMTC ticketing data - both from electronic ticketing and manual ticketing, is available for 4 depots. The ticketing data is available for depot - 6,25,28 and 41. The columns available in ticketing data are:

Ticket\_id, waybil\_Id, waybill\_no, schedule\_no, trip\_no, etim\_noroute\_no, route\_id, transaction\_no, ticket\_no, ticket\_type\_short\_code, ticket\_sub\_type\_short\_code, ticket\_from\_stop\_id, ticket\_from\_stop\_seq\_no, ticket\_till\_stop\_id, fare\_type, upload\_flag, ticket\_code, bus\_service\_id, epurse\_last\_from\_stop\_code, ticket\_printed\_flag, epurse\_last\_to\_stop\_id, epurse\_last\_ticket\_time, inserted\_by, inserted\_date, epurse\_last\_waybill\_no\_qstr, vehicle\_no, pass\_id\_no, depot\_id, epurse\_card\_no, epurse\_card\_issue\_name, px\_total\_amount, ticket\_date, ticket\_time, payment\_mode, luggage\_total\_amount, concession\_type, ticket\_till\_stop\_seq\_no, total\_ticket\_amount, rgroup\_ticket\_mode, toll\_amount, epurse\_current\_amount\_qstr, epurse\_last\_amount\_qstr, epurse\_last\_ticket\_no, travelled\_KM, luggage\_count, px\_count, epurse\_last\_ticket\_date, trip\_status, trip\_end\_condition, waybill\_type, form\_four\_Id, shift\_no, conductor\_token\_id, rate\_id, stage\_count, service\_tax\_amt, stage\_ticket.

Since we are interested only in some specific columns like : **ticket\_till\_stop\_id, schedule\_no, trip\_no, vehicle\_no, ticket\_type\_short\_code, px\_count, total\_ticket\_amount, ticket\_sub\_type\_short\_code, ticket\_from\_stop\_seq\_no, ticket\_printed\_flag, ticket\_date, ticket\_time, route\_no, etim\_no, shift\_no, ticket\_till\_stop\_seq\_no, ticket\_from\_stop\_id, trip\_direc.**

***This data has been extracted route-wise from the given dump and is available readily.***

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **Ticket\_from\_stop\_id** | Origin stop for the ticket sold |
| **Ticket\_till\_stop\_id** | Destination stop for the ticket sold |
| **Schedule\_no** | Bus Schedule number |
| **Trip\_no** | Trip number of the schedule |
| **Vehicle\_no** | Vehicle number of the bus [KA01F9372 ] |
| **Ticket\_type\_short\_code** | Code for type of ticket sold [Trip start / Trip close / Passenger / Luggage / Group / Pass / Penalty / Stage close / Toll pass etc] |
| **Ticket\_sub\_type\_short\_code** | Subtype of the ticket sold like [Adult / Child / Heavyweight / Lightweight / Daily Pass etc] |
| **px\_count** | Number of passengers |
| **Total\_ticket\_amount** | Amount of the ticket sold |
| **Ticket\_from\_stop\_seq\_no** | Within the route, stop\_no from where the passenger boards the bus |
| **Ticket\_till\_stop\_seq\_no** | Within the route, stop\_no where the passenger alights from the bus |
| **Ticket\_printed\_flag** | Whether the ticket was printed |
| **Ticket\_date** | Ticket issue date |
| **Ticket\_time** | Ticket issue time |
| **Route\_no** | Route number as displayed in the bus |
| **Etim\_no** | Electronic ticketing machine number |
| **Shift\_no** | Shift number for the given trip under the given schedule number |
| **Trip\_direc** | Trip direction whether it is forward(UP) or backward(DN) |

**Example Data:**

**ticket\_till\_stop\_id schedule\_no trip\_no vehicle\_no ticket\_type\_short\_code px\_count total\_ticket\_amount ticket\_sub\_type\_short\_code ticket\_from\_stop\_seq\_no ticket\_printed\_flag ticket\_date\_time route\_no etim\_no shift\_no ticket\_till\_stop\_seq\_no ticket\_from\_stop\_id trip\_direc**

0 138/4 3 KA01F9372 0 0 0 TS 1 N 2018-12-03 06:54:18 138 325-738-312 4 0 0 DN

1632 138/4 3 KA01F9372 1 1 5 AD 11 Y 2018-12-03 06:56:45 138 325-738-312 4 15 476 DN

8456 138/4 3 KA01F9372 1 1 17 AD 11 Y 2018-12-03 06:57:08 138 325-738-312 4 23 476 DN

1632 138/4 3 KA01F9372 1 1 5 AD 11 Y 2018-12-03 06:57:25 138 325-738-312 4 15 476 DN

1632 138/4 3 KA01F9372 1 1 5 AD 11 Y 2018-12-03 06:57:45 138 325-738-312 4 15 476 DN

8456 138/4 3 KA01F9372 1 1 17 AD 11 Y 2018-12-03 07:02:43 138 325-738-312 4 23 476 DN

1534 138/4 3 KA01F9372 1 1 10 AD 11 Y 2018-12-03 07:02:58 138 325-738-312 4 16 476 DN

8456 138/4 3 KA01F9372 4 1 0 MP 11 Y 2018-12-03 07:04:53 138 325-738-312 4 23 476 DN

8456 138/4 3 KA01F9372 4 1 0 DP 11 Y 2018-12-03 07:05:02 138 325-738-312 4 23 476 DN

8456 138/4 3 KA01F9372 4 1 0 DP 11 Y 2018-12-03 07:05:08 138 325-738-312 4 23 476 DN

8456 138/4 3 KA01F9372 4 1 0 DP 11 Y 2018-12-03 07:05:14 138 325-738-312 4 23 476 DN

8456 138/4 3 KA01F9372 4 1 0 MP 15 Y 2018-12-03 07:05:34 138 325-738-312 4 23 1632 DN

8456 138/4 3 KA01F9372 4 1 0 MP 15 Y 2018-12-03 07:06:23 138 325-738-312 4 23 1632 DN

8456 138/4 3 KA01F9372 4 1 0 DP 15 Y 2018-12-03 07:06:35 138 325-738-312 4 23 1632 DN

8456 138/4 3 KA01F9372 4 1 0 DP 15 Y 2018-12-03 07:06:49 138 325-738-312 4 23 1632 DN

8456 138/4 3 KA01F9372 4 1 0 MP 15 Y 2018-12-03 07:07:02 138 325-738-312 4 23 1632 DN

8456 138/4 3 KA01F9372 4 1 0 MP 15 Y 2018-12-03 07:07:13 138 325-738-312 4 23 1632 DN

8456 138/4 3 KA01F9372 1 2 10 AD 18 Y 2018-12-03 07:13:32 138 325-738-312 4 23 8379 DN

0 138/4 3 KA01F9372 0 0 0 TL 23 N 2018-12-03 07:22:03 138 325-738-312 4 0 0 DN

**Scope or work:**

**Data Provided by BMTC**

The raw data as a single mysql server dump is provided to us from BMTC. The size of the dump file differs between 5 GB to 50 GB depending on the number of months and numbers of depots for which the data is provided.

In our case, we are provided with one month ticket sales data for December 2018 data from the depots : 6, 25,28 and 41 for all the routes that operate out of these depots. Other supplementary data with respect to:

1. route information (route.csv, route\_point.csv),
2. schedule information(11\_10\_form\_four\_Jan, 2019.xlsx),
3. ticket type information (ticket\_type.csv, ticket\_sub\_type.csv).

In addition to the ticket sales data are provided.

**Pre-processing of Data**

The ticket sales data that is provided as a single dump file needs to be processed so that they can be used in any analytic tasks. Since the data was huge, we need to break it into manageable chunks with which we can work. Since, we have been using Python for all the analytic tasks we needed the data to be in different format other than mysql dump. Hence, the raw data was converted into manageable smaller sized .csv files.

From the raw ticket sales data in .csv files, the route wise ticket sales data was generated for every depot.

The generated route wise ticket sales data was sorted based on the following parameters in the order given: 'vehicle\_no','schedule\_no','shift\_no','trip\_no','ticket\_date\_time'.

Combine the same route\_no files from different depots into one single file for every route.

**Development of KPI’s**

Develop the following KPIs from Ticketing Data:

1. **Trip-wise Occupancy for all routes.**
   1. Compute the occupancy for every route along different stops of the route. The occupancy needs to be computed at every bus stop along the route. Simultaneously, the time at which the bus arrived at the intermediate stops also must be identified and stored from the ticket sales data.
   2. While computing the occupancy, simultaneously identify the different trips from the ticket sales data. Hence, the trip wise occupancy is computed from the ticket sales data for every route.
   3. The timestamps for all the intermediate bus stops might not be available in the ticket sales data. Hence, they need to be interpolated.
   4. Generate **real time occupancy <<route\_no>\_occ.csv>** for every trip with the following information.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Sample Value** |
| **trip\_date\_vehicle\_no** | Concatenated value of - Date of the trip\_ Vehicle registration number | 2018-12-29\_ KA01FA1010 |
| **trip\_no** | Trip number as obeserved from the data | 8 |
| **trip\_direction\_busstop\_name** | Concatenated values of Trip direction – UP or DOWN(DN) with **'\_’** and bus stop name | up\_Vijayanagara |
| **board** | Number of people boarded at the stop in that trip | Any whole number |
| **alight** | Number of people got down (alighted) at the stop in that trip | Any whole number |
| **occupancy** | Number of people who are currently inside the bus at a stop after boarding and alighting | Any whole number |
| **occ\_time** | Time stamp of ticket sales at this stop | Date\_time (2018-12-29 08:20:48) |

1. **Actual number of trips:**
   1. Using the occupancy csv generated in previous step, generate one more file called **<<route\_no>\_realtrips.csv>** with the following information

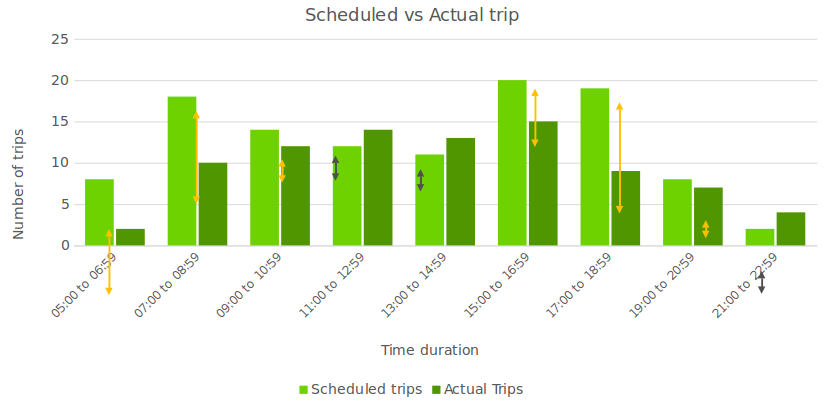
|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Sample Value** |
| **Sno** | Row id (Sequential numbering for the rows) | 1 |
| **trip\_id** | Value obtained by concatenating the following columns : **Sno, trip\_date, vehicle\_no, trip\_no, trip\_direction \_busstop\_name** | 2018-12-29\_KA01FA1010\_8\_DN |
| **from\_bus\_stop\_id** | Stop id of Origin as in ticket sales | 1 |
| **to\_bus\_stop\_id** | Stop\_id of Destination as in ticket sales | 32 |
| **observed\_trip\_start\_time** | Observed Time stamp at the start of the trip or time stamp from the first ticket sale. | 6:14:24 |
| **observed\_trip\_end\_time** | Observed Time stamp at the start of the trip or time stamp from the last ticket sale. | 7:52:58 |
| **travel\_duration** | Difference between **observed\_trip\_end\_time** and **observed\_trip\_start\_time** | 1:11:34 |
| **computed\_speed** | Speed to be computed using the formula: 22/(HOUR(**travel\_duration**)+(MINUTE(**travel\_duration**)/60)) | 18.59 |
| **whether\_boardings\_observed** | Was there any ticket sales in the identified trip: **Yes** or **No** | Yes |
| **trip\_start\_hour\_of\_day** | The hour of the day in which the identified trip started from its origin | 7 |

* 1. From the **<<route\_no>\_realtrips.csv>** file count the number of trips with boarding=’**yes’ for every hour of the day separately.** This gives the number of actual trips for every route number.

1. **Planned number of trips:**
   1. Use the form 4 data in (**11\_10\_form\_four\_Jan, 2019.xlsx**) to generate the number of planned trips. For every schedule in form4 with only trips of Origin-Destination (as in the **<<route\_no>\_realtrips.csv>**) sheet for a given route, count the number of trips for every hour of the day separately. This gives the number of planned trips.
   2. Generate the file **<trip\_kpi.csv>** with the following information.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Sample Value** |
| **Sno** | Row id (Sequential numbering for the rows) | 1 |
| **trip\_id** | Value obtained by concatenating the following columns : **Sno, trip\_date, vehicle\_no, trip\_no, trip\_direction \_busstop\_name** | 2018-12-29\_KA01FA1010\_8\_DN |
| **hour\_of\_day** | Hour of the day |  |
| **trip\_direction** | Trip direction |  |
| **no\_actual\_trips** | Number of trips as counted for that hour of the day and in that direction(up/dn) from **step(f).** |  |
| **no\_planned\_trips** | Number of trips as counter for that hour of the day and in that direction(up/dn) from **step(g)** |  |

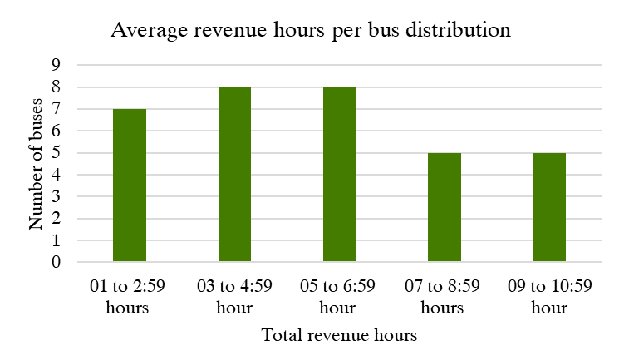
* 1. Generate bar graph / histogram using the file **<trip\_kpi.csv>**.



1. **Revenue hours per bus:**

Use the following columns from **<<route\_no>\_realtrips.csv>** file : Route\_no, vehicle\_no, travel\_duration .

* 1. Get the travel duration from every trip, every bus and every date is obtained.
  2. Average of travel\_duration = Average revenue hours of all buses
  3. Average of travel time by on bus across different days = Average revenue hour per bus per day



1. **Dispatch Inconsistency:** 
   1. For every route, count the number of vehicle that are available for every hour on different days of the month UP/DN separeately.
   2. Generate a table as shown below:

