**Background Study**

In every IT company, there is help desk unit which provides the facilities to raise ticket if any technical or non-technical issues occurs or incident take place occurs during day to day operations of the organization. The issues can be impacting a single person, a team, a department or the entire organization. By raising ticket, the user can take assistance from the help desk and the help desk representatives in turn solve the issues depending on the level of priority of the incident.

Now the priority depends on the two things,

1. Impact
2. Urgency
3. **Impact**: - A measure of the effect of an Incident on organizational processes. It measures the number of clients potentially affected by an Incident.

It is categorized into following types:

|  |  |  |
| --- | --- | --- |
| **Sl No.** | **Impact** | **Description** |
| 1. | High | * A large number of staff are affected and/or not able to do their job. * A large number of customers are affected and/or acutely disadvantaged in some way. * The financial impact of the Incident is (for example) likely to exceed $10,000. * The damage to the reputation of the business is likely to be high. * Someone has been injured. |
| 2. | Medium | * A moderate number of staff are affected and/or not able to do their job properly. * A moderate number of customers are affected and/or inconvenienced in some way. * The financial impact of the Incident is (for example) likely to exceed $1,000 but will not be more than $10,000. * The damage to the reputation of the business is likely to be moderate. |
| 3. | Low | * A minimal number of staff are affected and/or able to deliver an acceptable service but this requires extra effort. * A minimal number of customers are affected and/or inconvenienced but not in a significant way. * The financial impact of the Incident is (for example) likely to be less than $1,000. * The damage to the reputation of the business is likely to be minimal. |

2. **Urgency: -** It is measured by how quickly the incident needs to be resolved. It is a measure of how long it will be until an Incident has a significant Impact on the organization.

Eg:- A high Impact Incident may have low Urgency, if the Impact will not affect the organization until the end of the financial year.

High Incident causes immediate and significant disruption but not affecting life, safety, business transaction-critical and teaching-related services while in use.

Moderate Incident will cause disruption in the near term.

It is categorized in the following ways

|  |  |  |
| --- | --- | --- |
| **Sl. No** | **Urgency** | **Description** |
| 1. | High | * The damage caused by the Incident increases rapidly. * Work that cannot be completed by staff is highly time sensitive. * A minor Incident can be prevented from becoming a major Incident by acting immediately. * Several users with VIP status are affected. |
| 2. | Medium | * The damage caused by the Incident increases considerably over time. * A single user with VIP status is affected. |
| 3. | Low | * The damage caused by the Incident only marginally increases over time. * Work that cannot be completed by staff is not time sensitive. |

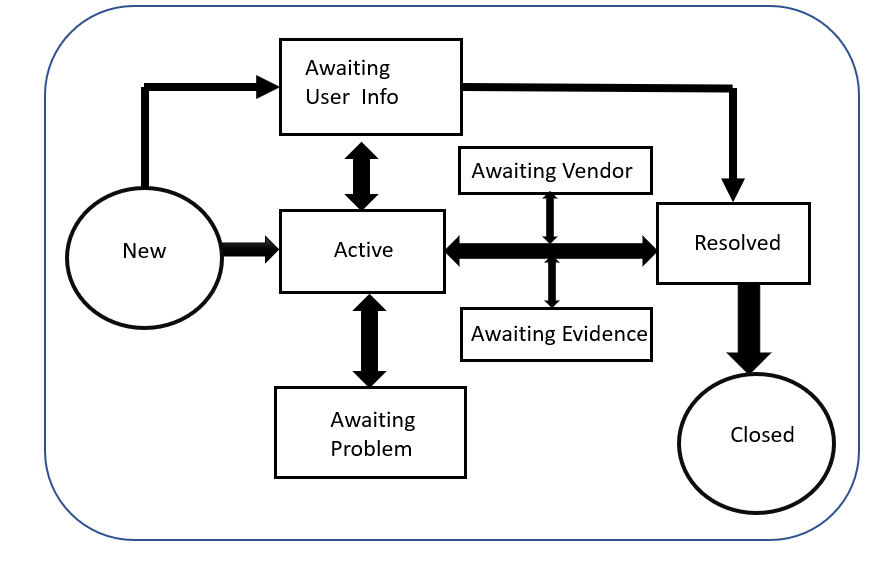
**Priority: - It** refers which task to solve first i.e., sequence of solving task

**Priority Matrix**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Priority** | | **Impact** | | |
| **High** | **Medium** | **Low** |
| **Urgency** | **High** | Critical | High | Moderate |
| **Medium** | High | Moderate | Low |
| **Low** | Moderate | Low | Low |

1. Low-priority incidents are those that do not interrupt users or the business and can be worked around. Services to users and customers can be maintained.
2. Medium-priority incidents affect a few staff and interrupt work to some degree. Customers may be slightly affected or inconvenienced.
3. High-priority incidents affect a large number of users or customers, interrupt business, and affect service delivery. These incidents almost always have a financial impact.

**Ticket Processing Flow**



**Business Objective:**

To predict the impact of the incident raised by the customer

**Datasets Details:**

There are 141712 number of records and 25 columns

**Target or Dependent Variable:** Impact ----> High, Medium, Low

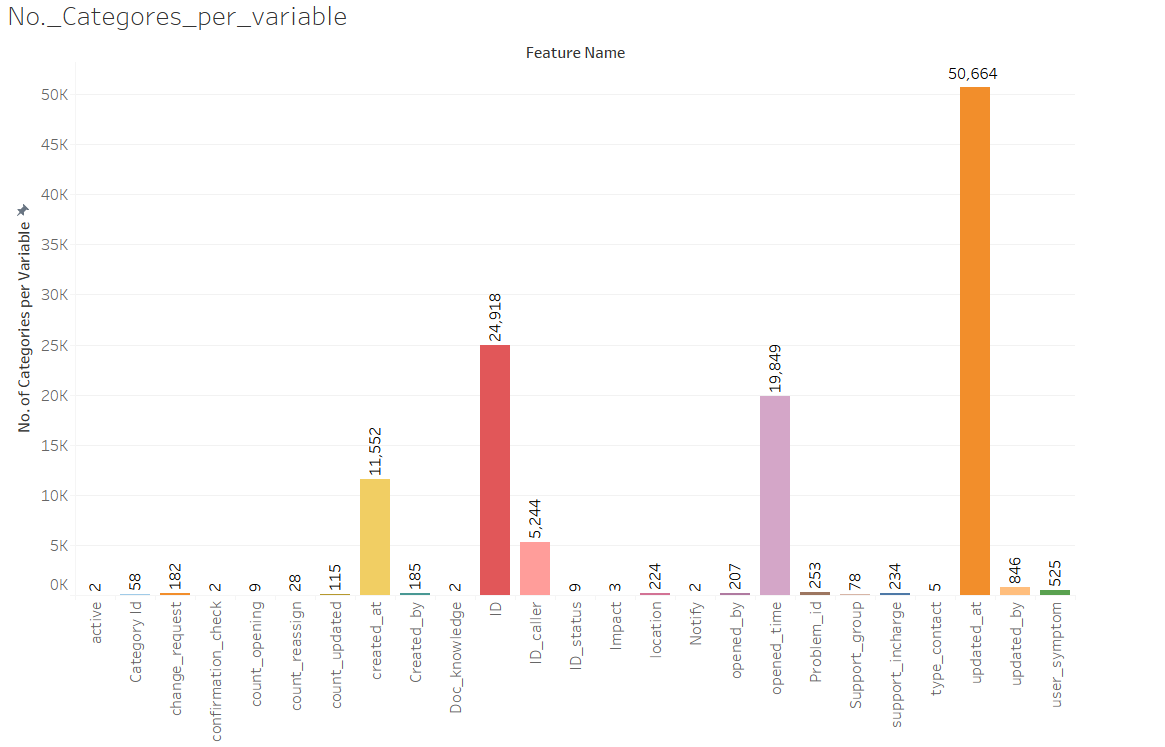
**Independent Variables:**

Categorical Variables:

|  |  |  |
| --- | --- | --- |
| Sl no. | Variables | Categories |
| 1. | ID | 24198 unique ID’s |
| 2. | ID\_status | -100, Active, Awaiting Evidence, Awating user info , Awaiting Problem, Awaiting Vendor, Closed, New, Resolved |
| 3. | active | True, False |
| 4. | ID\_caller | 5244 unique caller id’s |
| 5. | opened\_by | 207 unique users |
| 6 | created\_by | 185 unique identified |
| 7. | updated\_by | 846 unique users who updated the tickets |
| 8. | type\_contact | Direct opening, Email, IVR, Phone, Self service |
| 9. | location | 224 locations |
| 10 | category\_ID | 58 different categories |
| 11. | user\_symptom | 525 different user symptoms |
| 12. | Support\_group | 78 different support groups |
| 13. | support\_incharge | 234 unique support incharge |
| 14. | Doc\_knowledge | True, False |
| 15. | confirmation\_check | True, False |
| 16. | notify | Do Not Notify, Send Email |
| 17. | problem\_id | There are 253 different problems |
| 18. | change request | 182, different change request |

Quantitative Variables:

1. Numerical Variables
2. count\_reassign
3. count\_opening
4. count\_updated
5. Date/Time
6. Opened\_time
7. Created\_at
8. Updated\_at



**Exploratory Data analysis**

**Analysis of the variables which have influence in deciding the category of Impact**

1. ID Caller: - identifier of the user affected

This will identify the person who is affected by the incident. This can be useful in classify the impact.

Eg:- In an organization if large number of valuable client is getting affected because of the interrupted services impact can be high for that particular case.

1. Opened by: - Identifier of the user who reported the incident.

Incident can be reported by any associate who is link with the organization. Eg:- If the incident reported by any VIP level people it can be very important to solve immediately and can be categorized accordingly based on the impact.

1. Location: - identifier of the location of the place affected

This refers to the place which is affected by the service disruption. For some organization, certain location may be very important for business profit, if from such location incident is reported than it may cause high impact as it will be related to company’s finances

1. Category Id: - first-level description of the affected service

This will tell about the details about the service which is affected.

1. user\_symptom :- description of the user perception about service availability

This is what a user will describe what he is facing because of the incident.

1. Support\_group :- identifier of the support group in charge of the incident

This may have influence on the impact because support group may be different according to the type of incident raised. Different level of incidents will require different level of services, knowledge, resource technicians to resolve and they may belong to different support group.

1. support\_incharge: - identifier of the user in charge of the incident

Simmarly high critical incident may require different incharge compared to the incident with low impact.

1. confirmation\_check: - boolean attribute that shows whether the priority field has been double-checked.

This may indicate that service which is having high impact that may require high priority to solve immediately or at faster pace.

1. Problem\_id = identifier of the problem associated with the incident

This may say what problem arises because of the incidents

1. change\_request: - identifier of the change request associated with the incident

This can be important because this may identify what type of change is required for the incident raised.

**Feature Engineering**

One more variable may require which have to be created form the existing data

**Total\_time require to solve a particular problem**: - For example incident which have high impact requires less time to solve and vice versa.

Data Cleaning

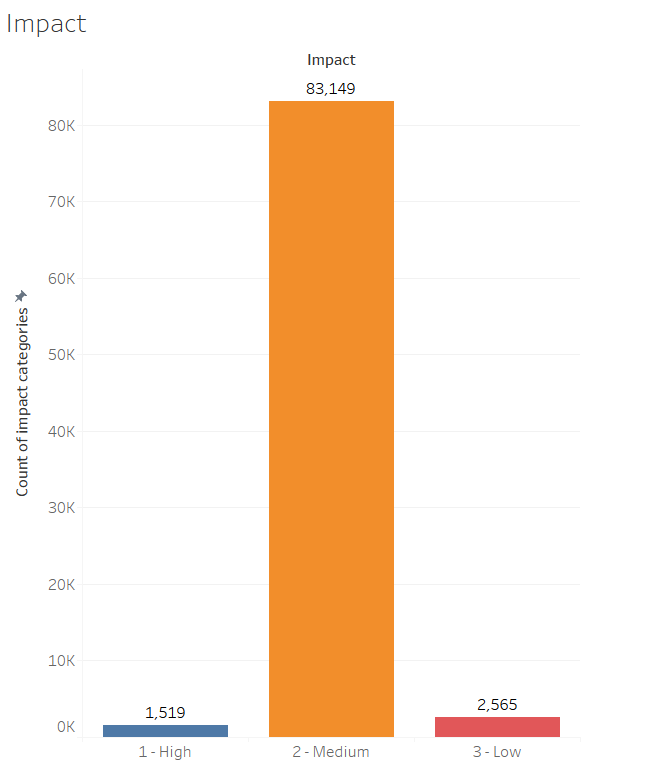
1. Sorting
2. Removing duplicate values

After removing duplicate values based on ID and ID status we get 87233 records

**Visualization**

1. **Impact**

Total unique Incidents = 24918 (based on our unique id’s)

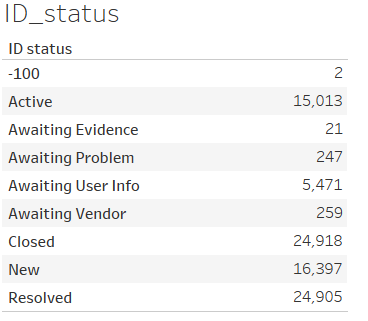
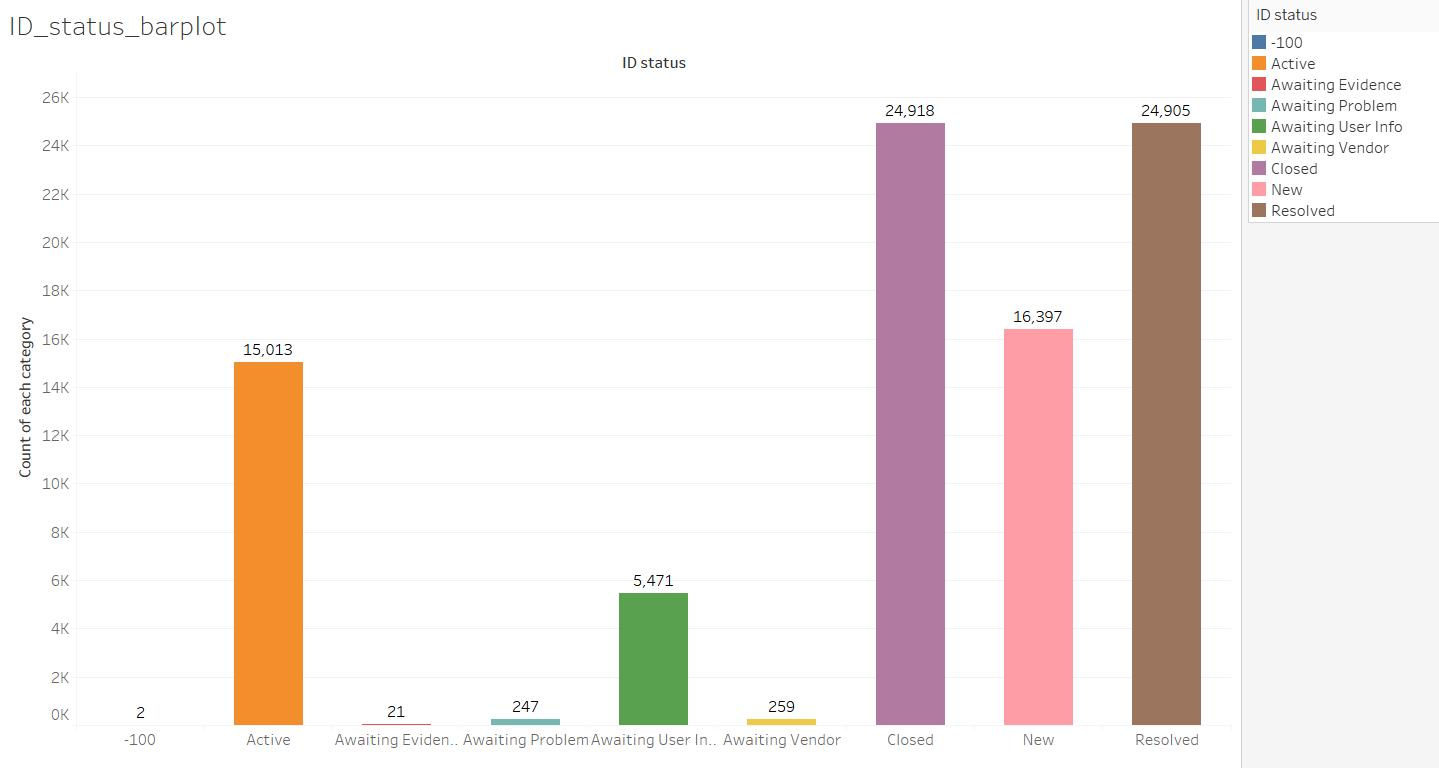


Out of 87233 total incidents, 1519 is having high impact, 83149 has medium impact and 2565 has low impact.

1. **ID**

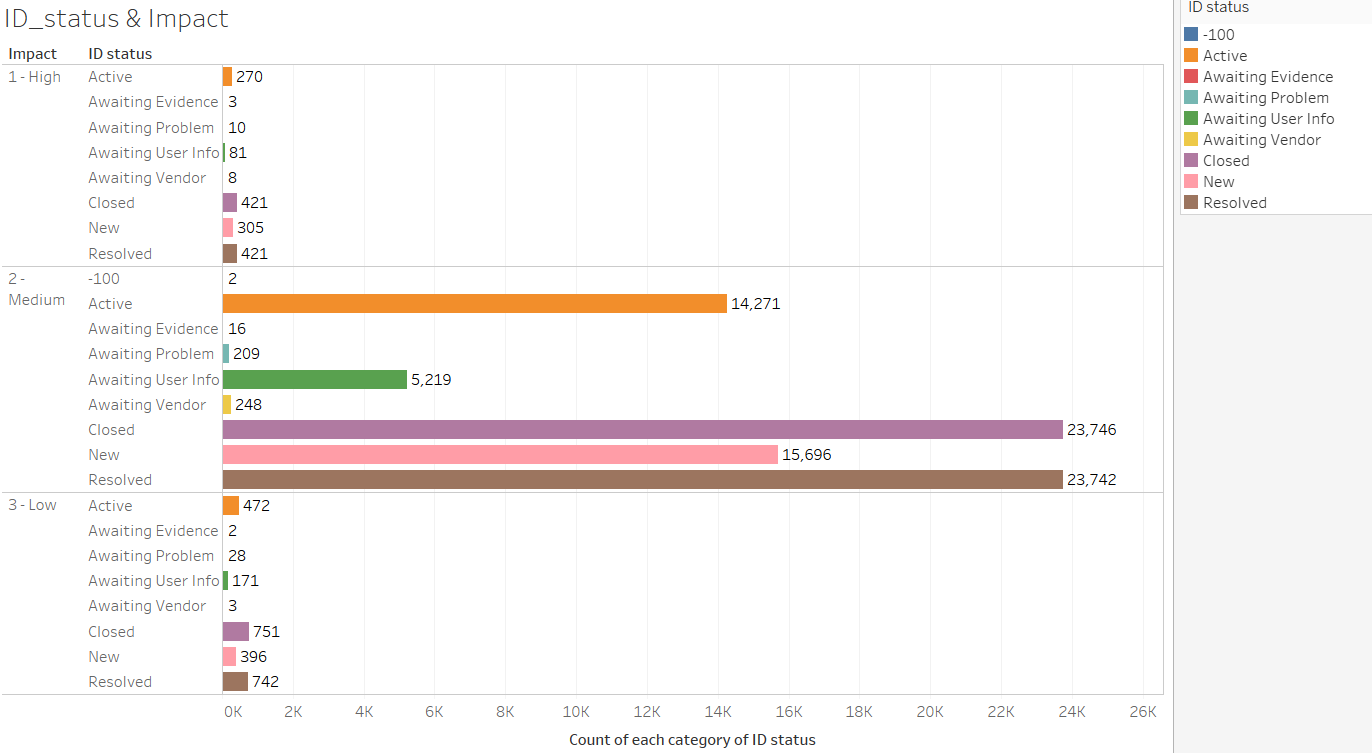
There are 24918 unique incident id’s are there in the given dataset.

1. Id\_Status

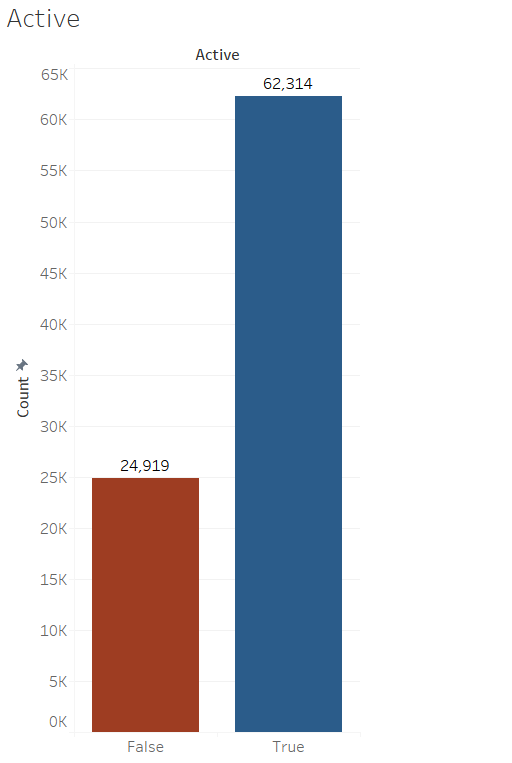
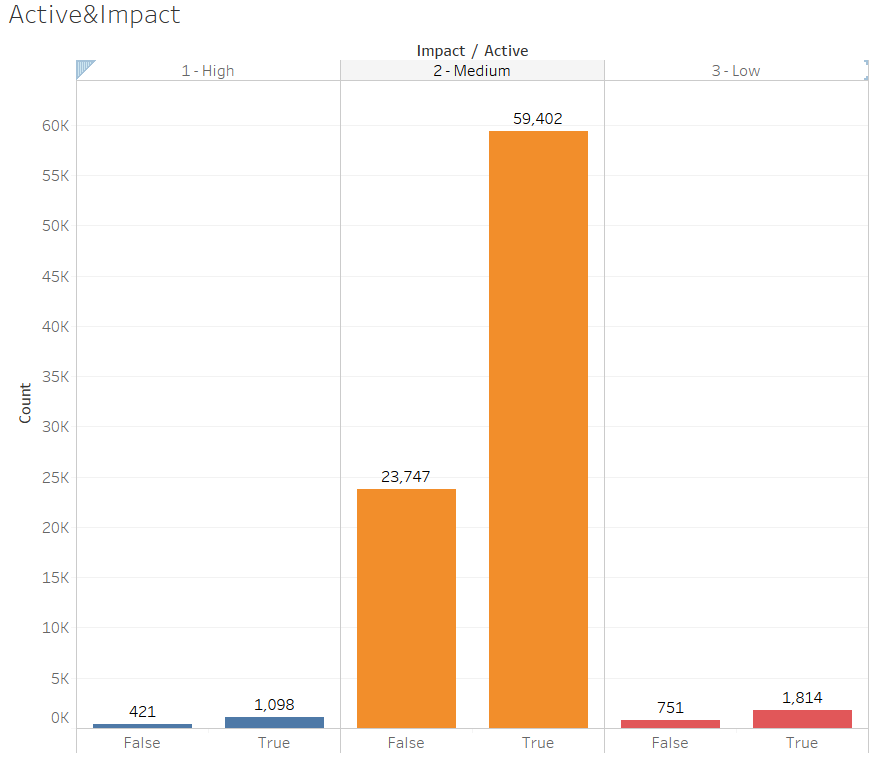
 

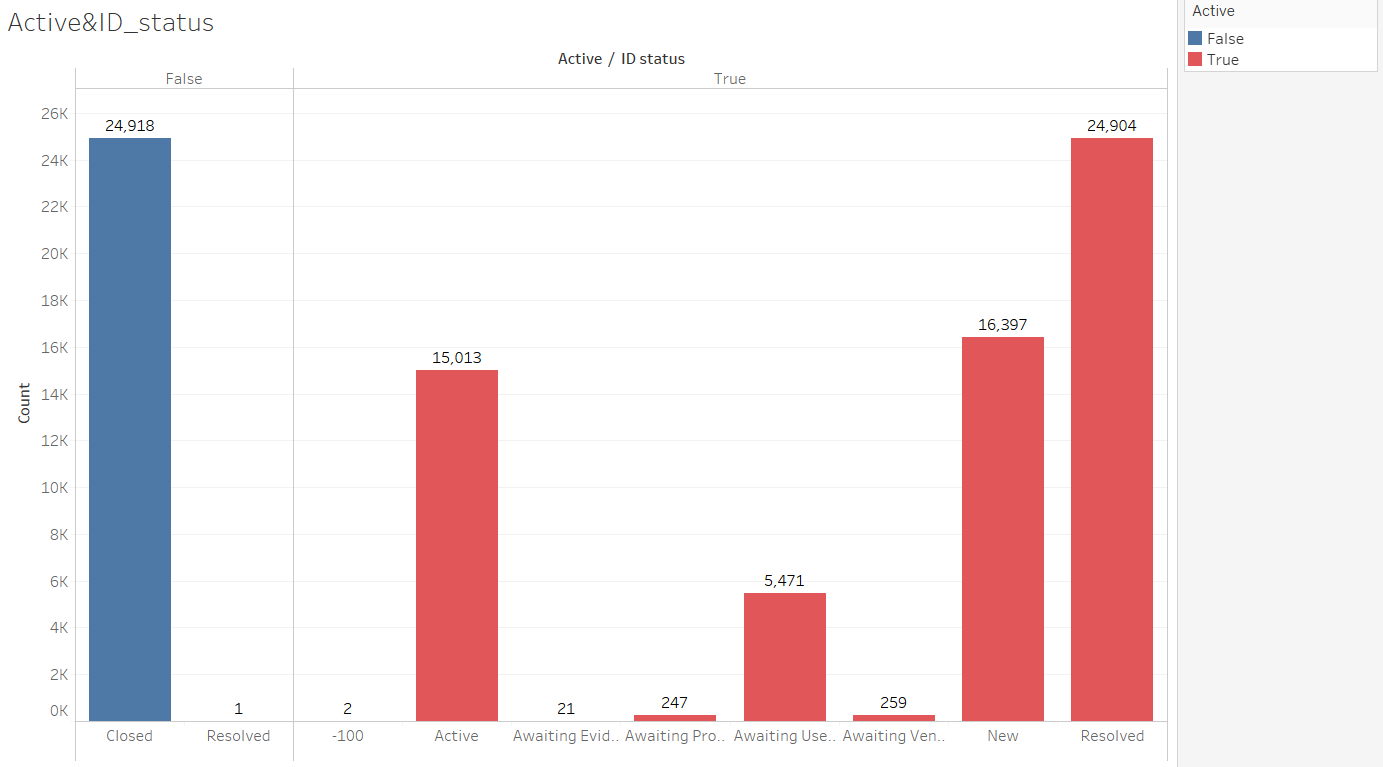
In the above polt, we can find that there is two records for -100 category, this can be an error in the dataset. This will be replaced with some other value or records can be deleted in further treatment

**Impact Vs ID status**



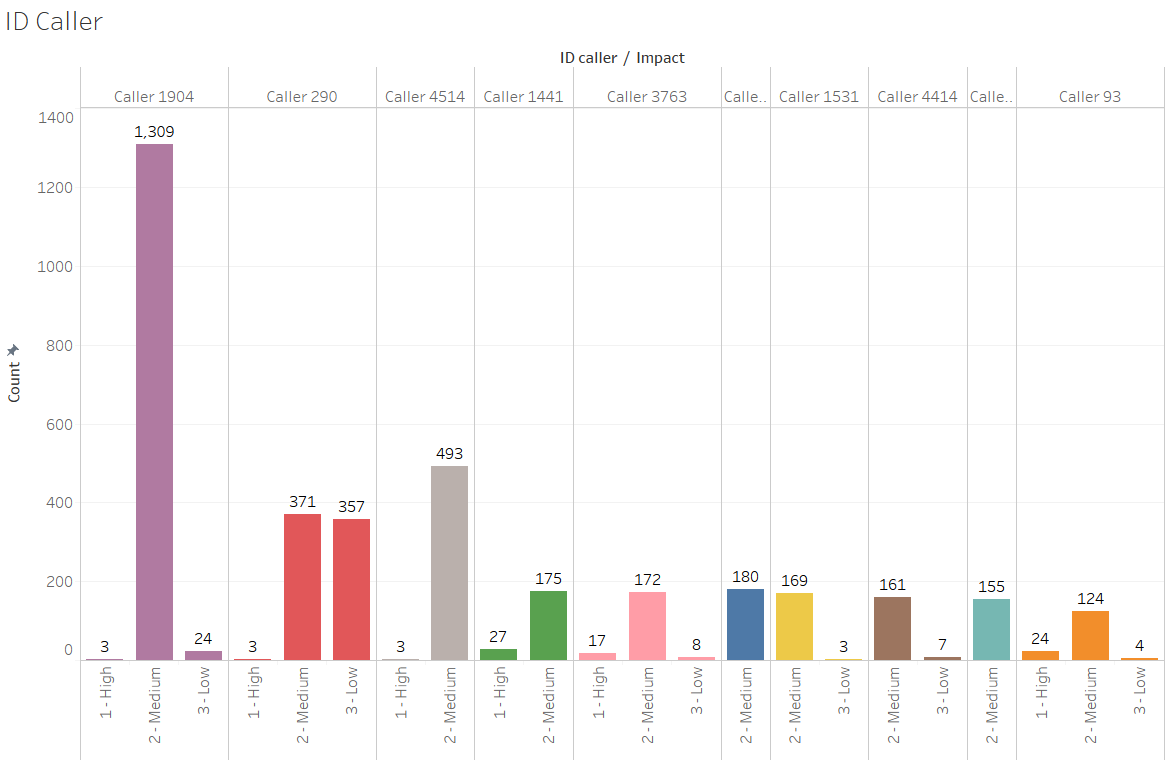
1. **Active**

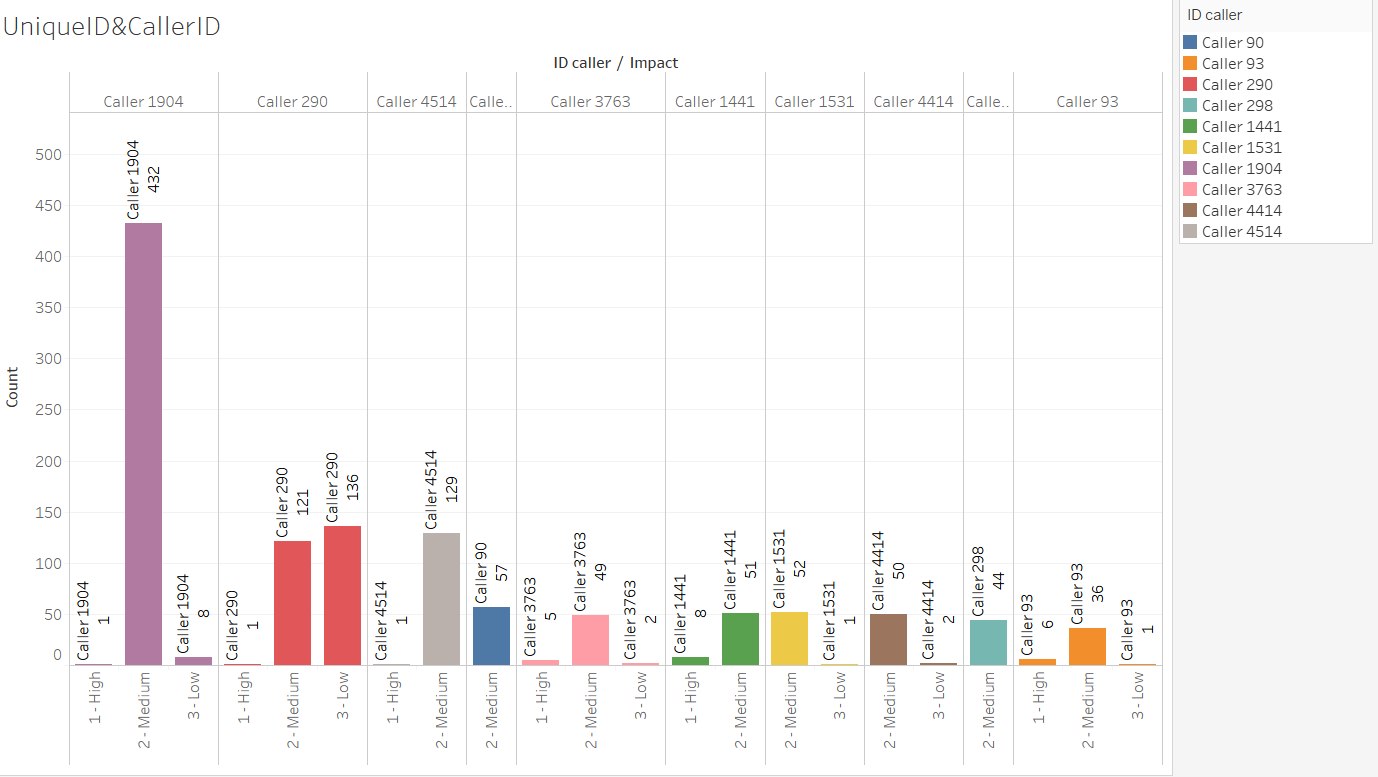
 



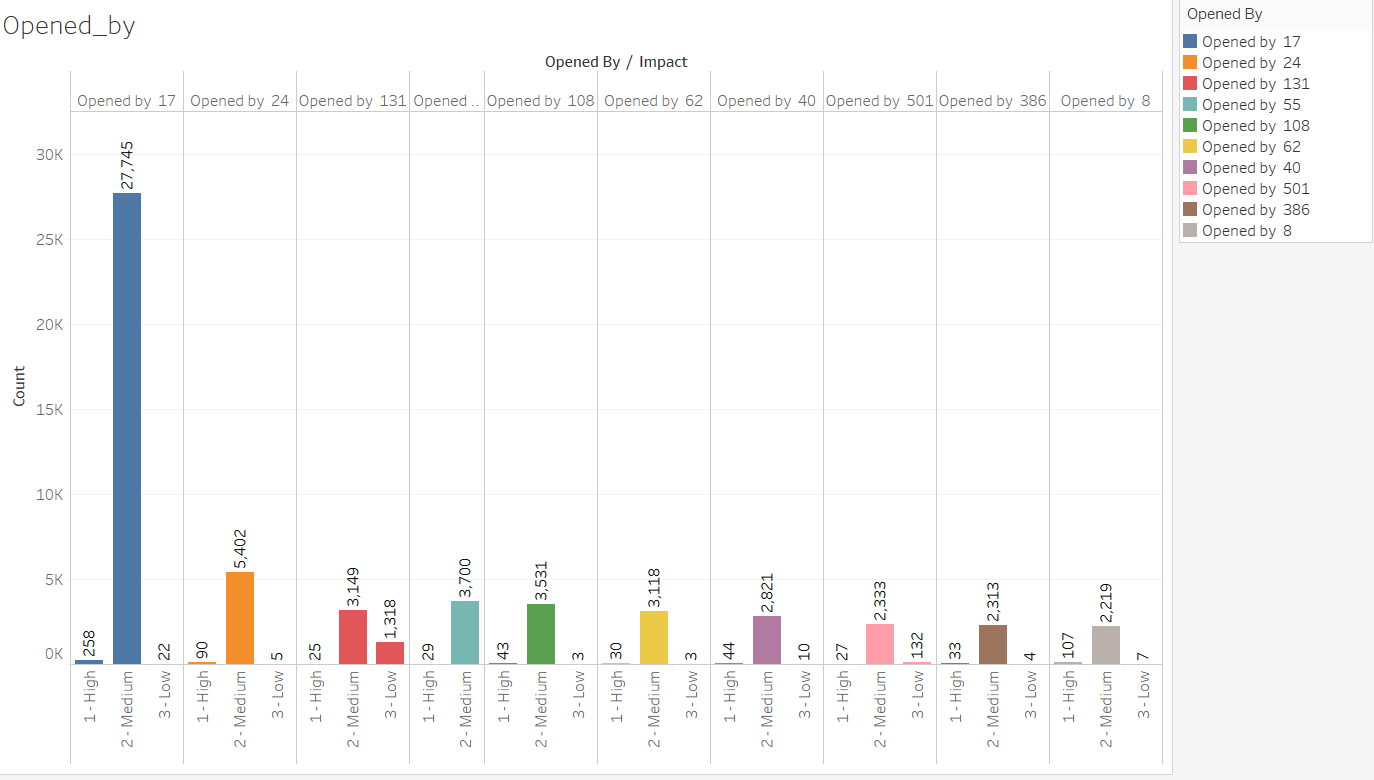
1. **ID\_Caller**

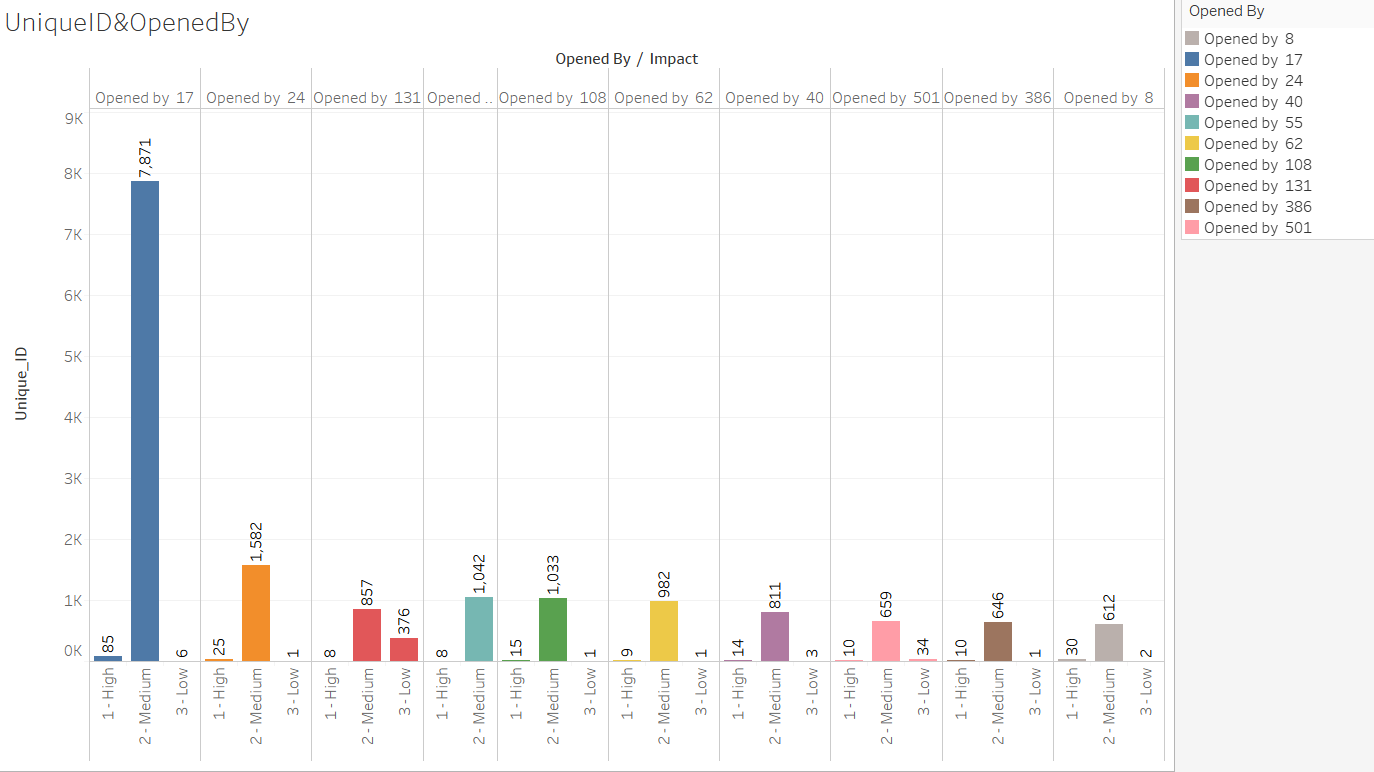
There are 5244 unique ID\_callers



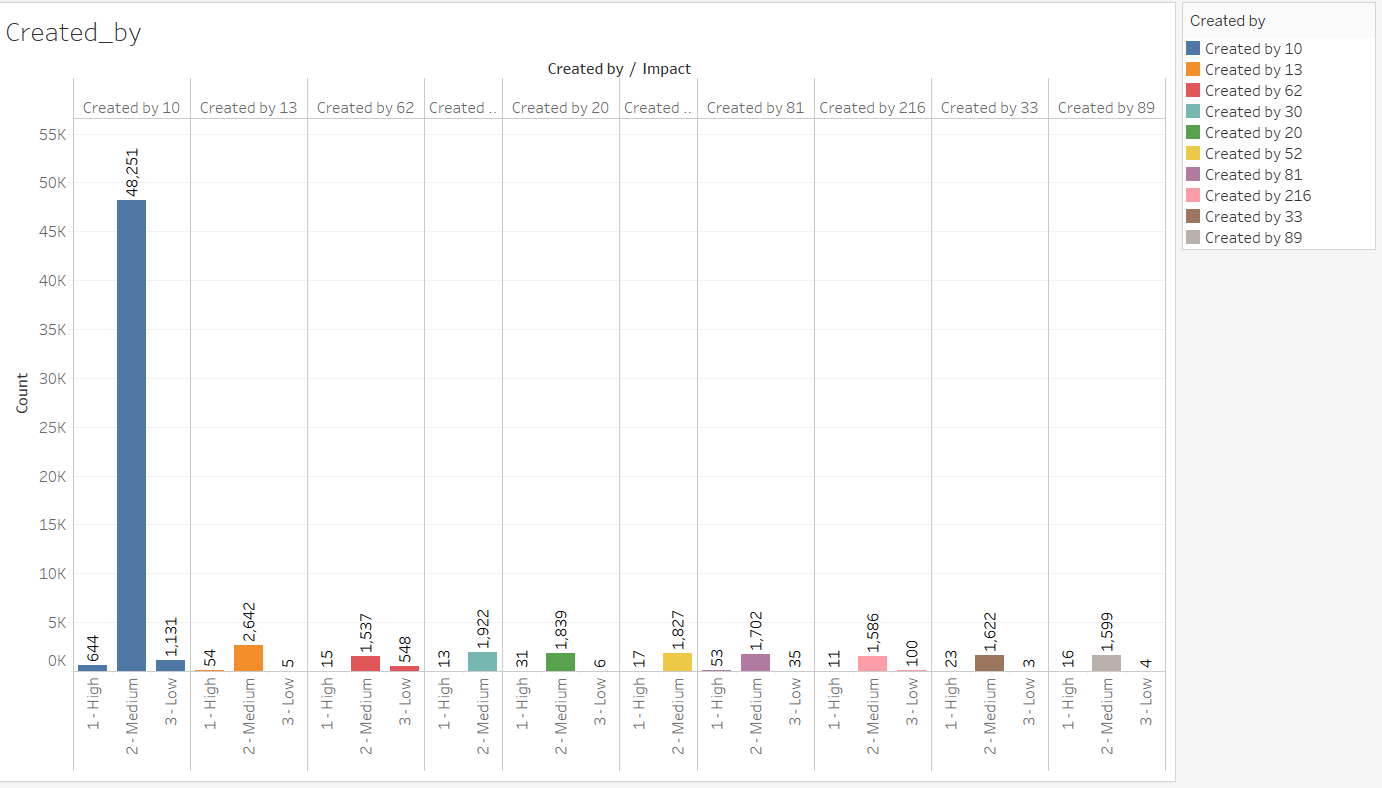


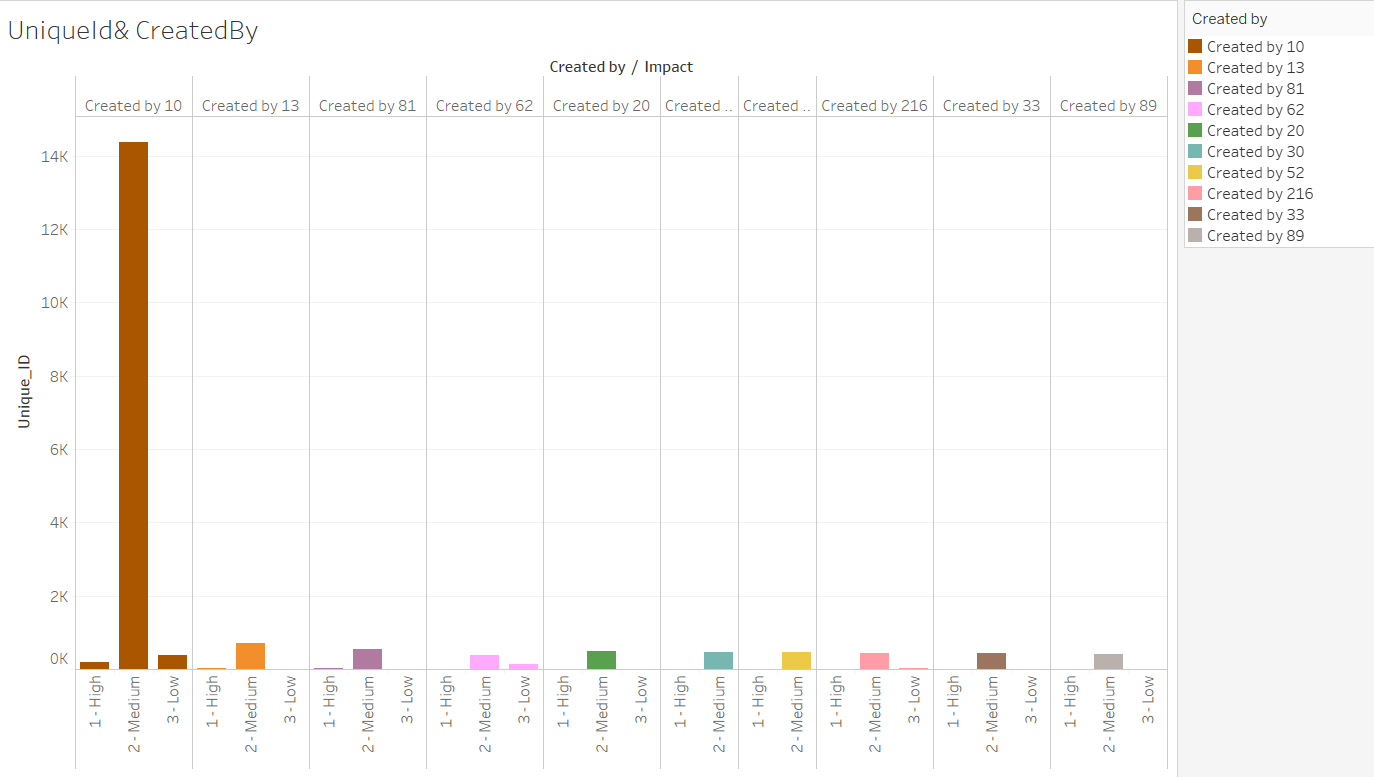
1. opened\_by 207 unique users



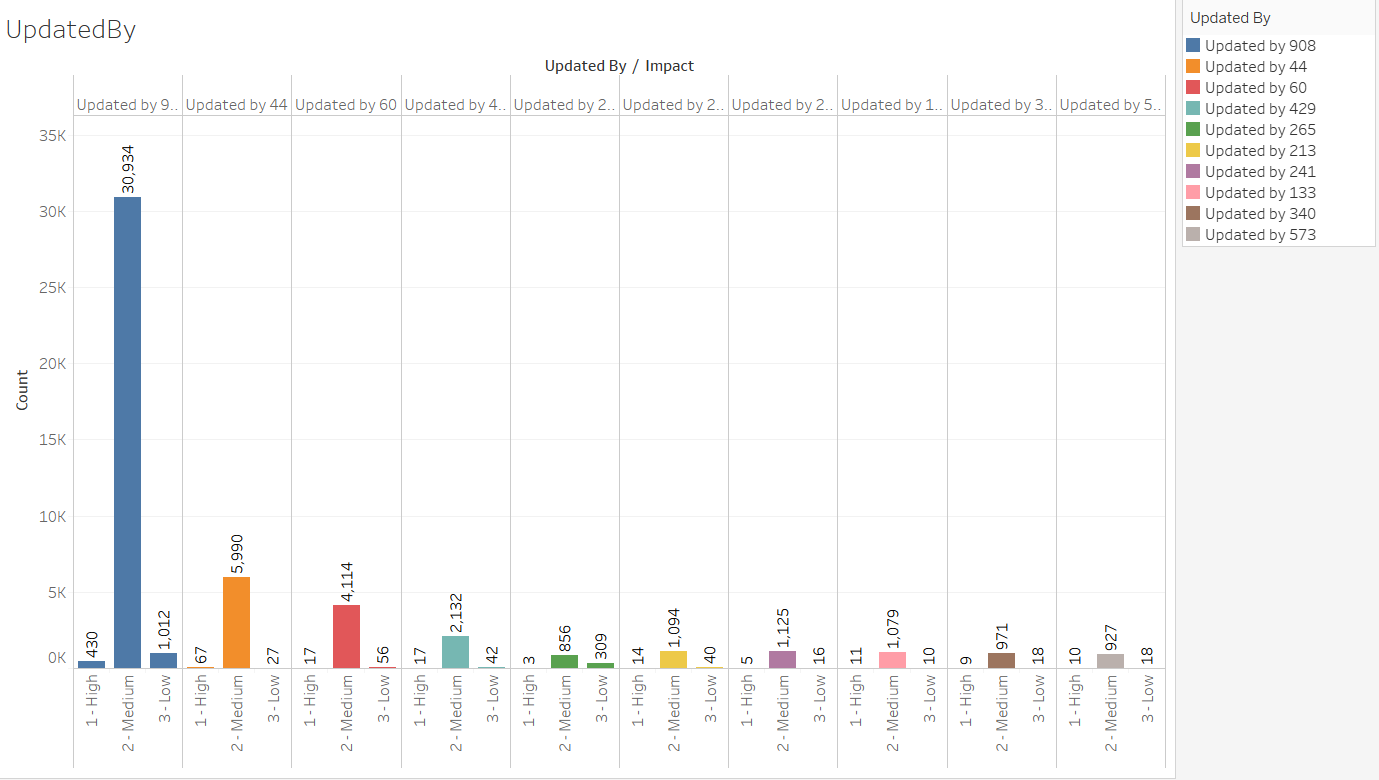


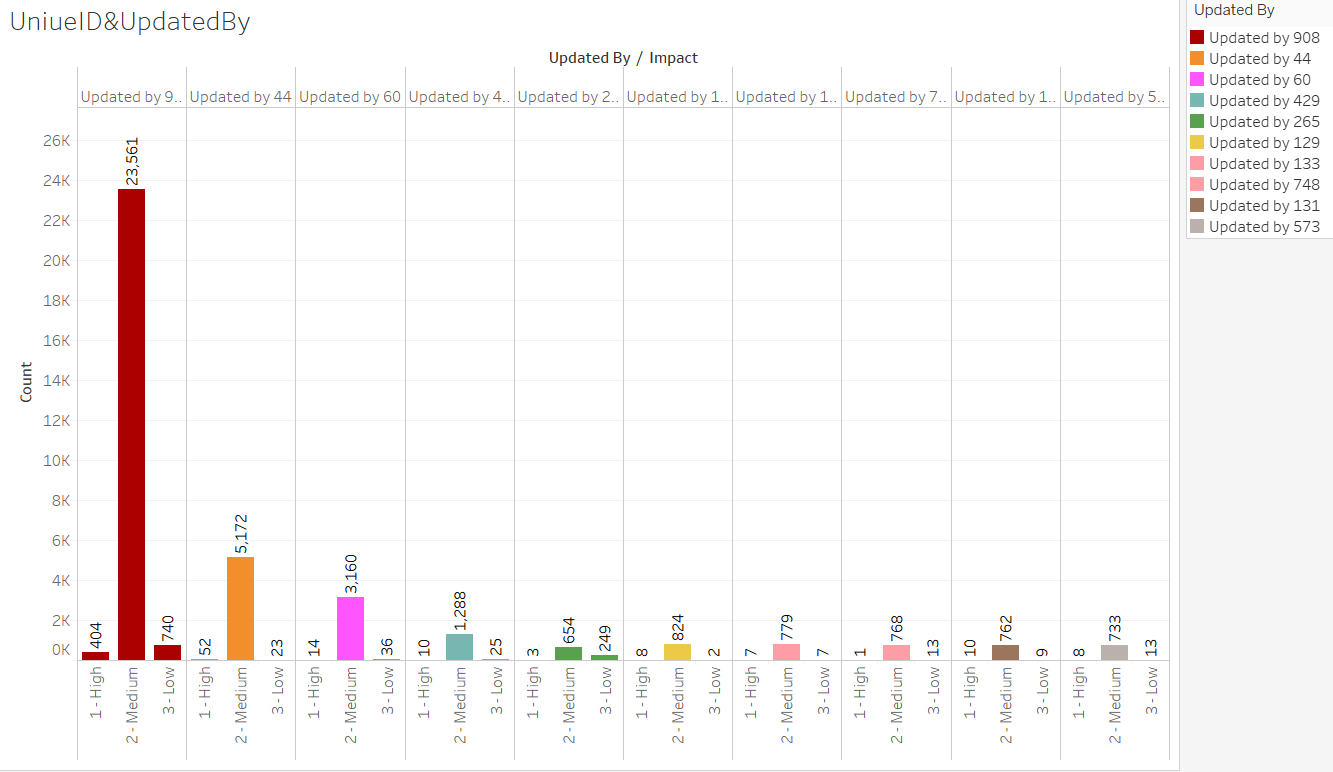
1. created\_by 185 unique identified



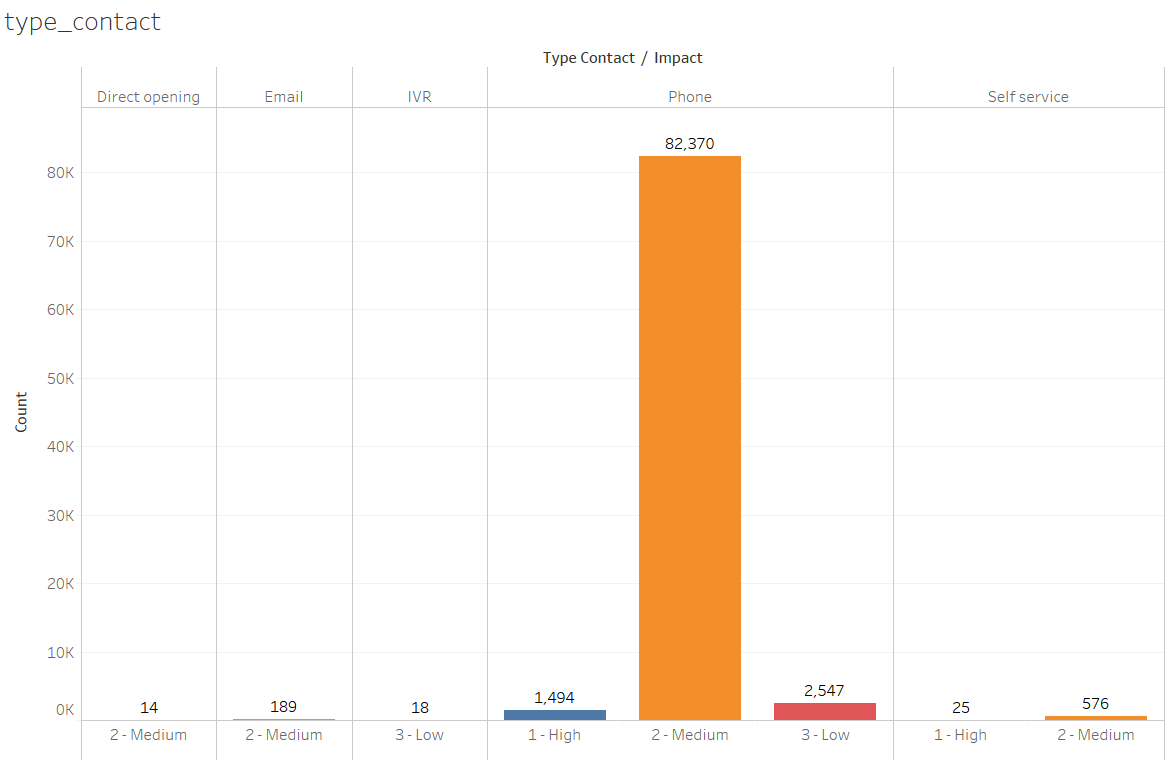
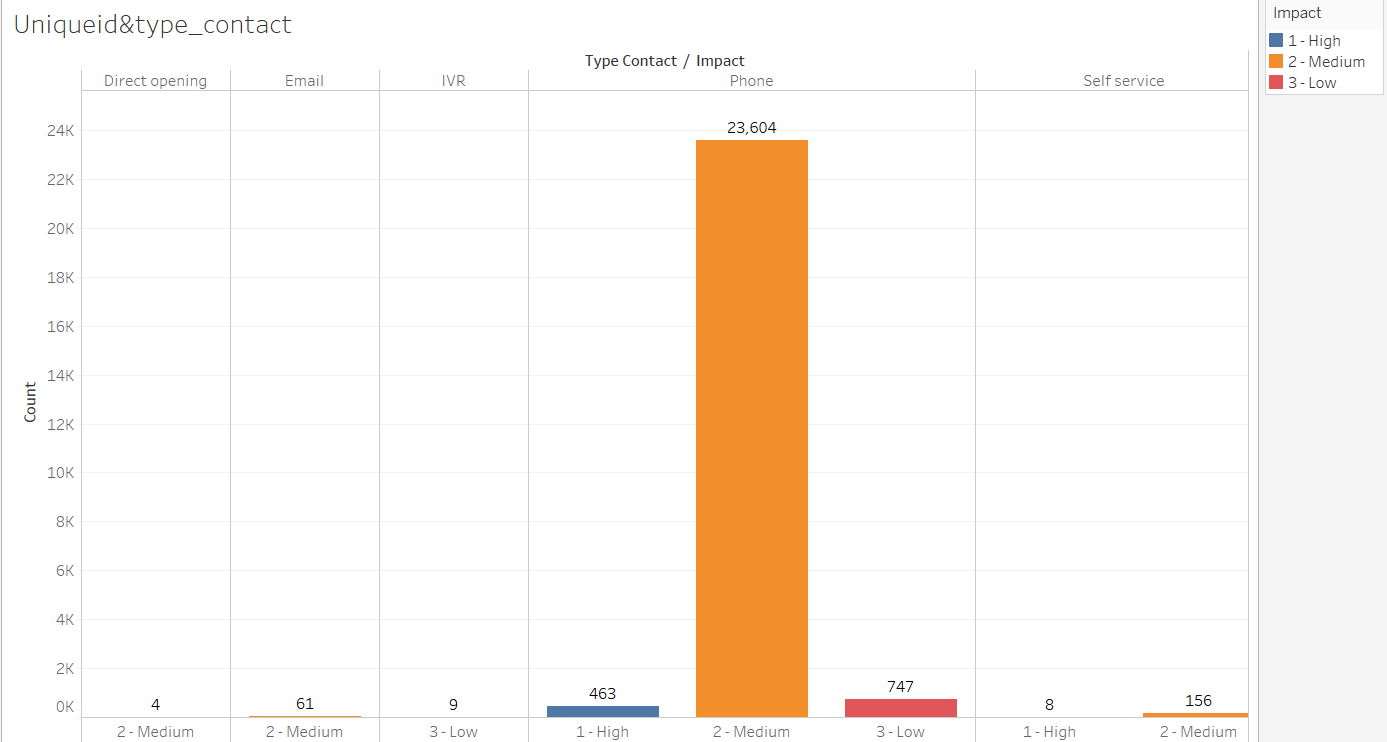


1. updated\_by 846 unique users who updated the tickets

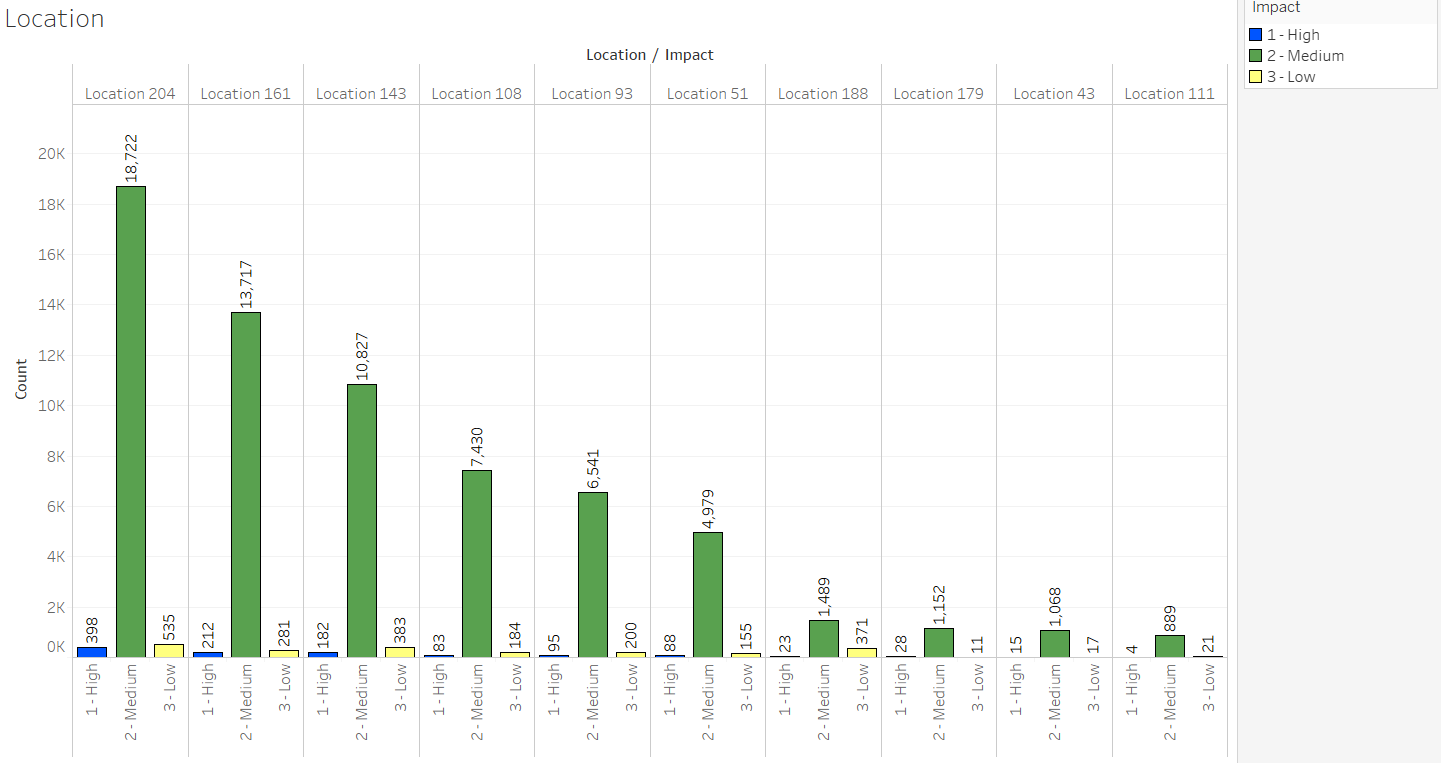


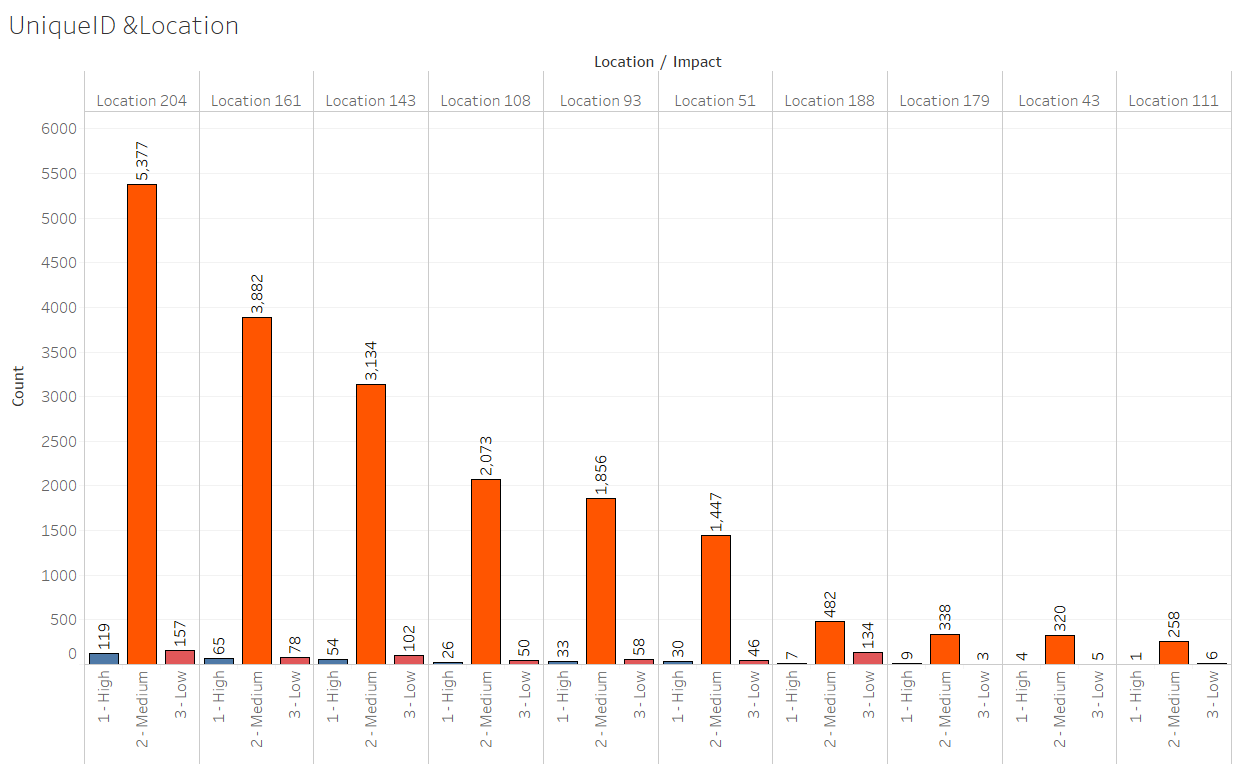


1. type\_contact-Direct opening, Email, IVR, Phone, Self service

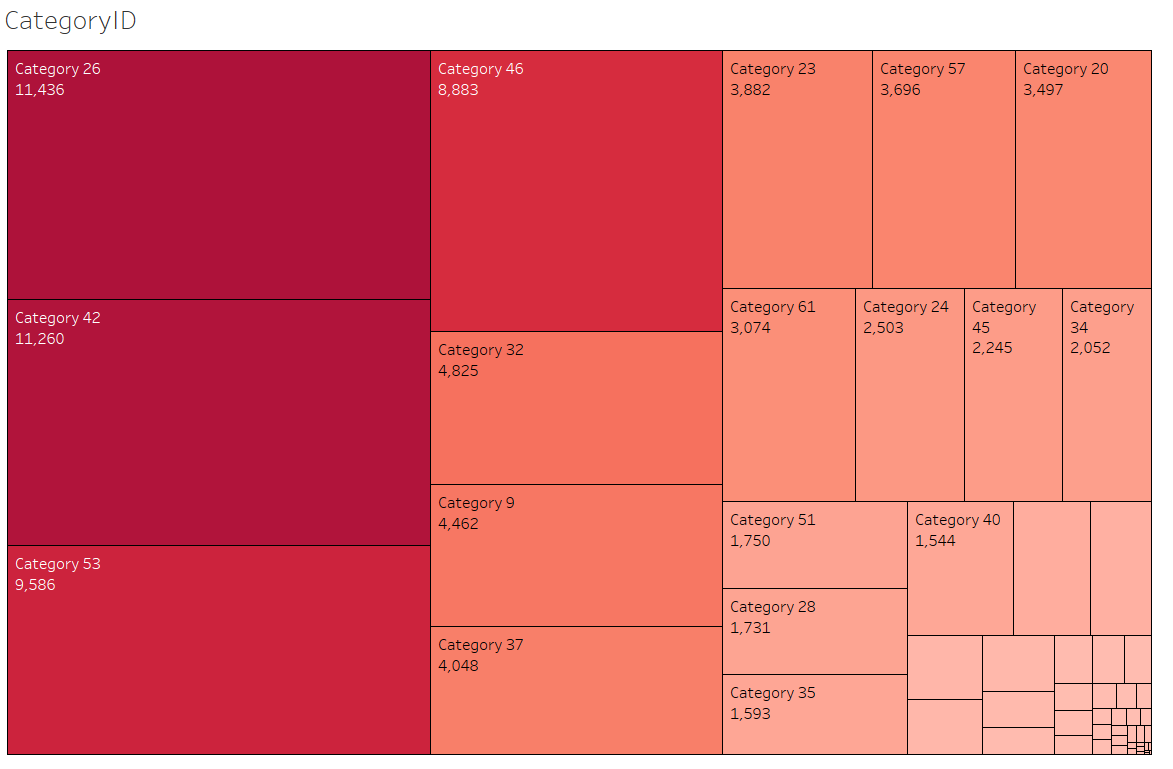
 

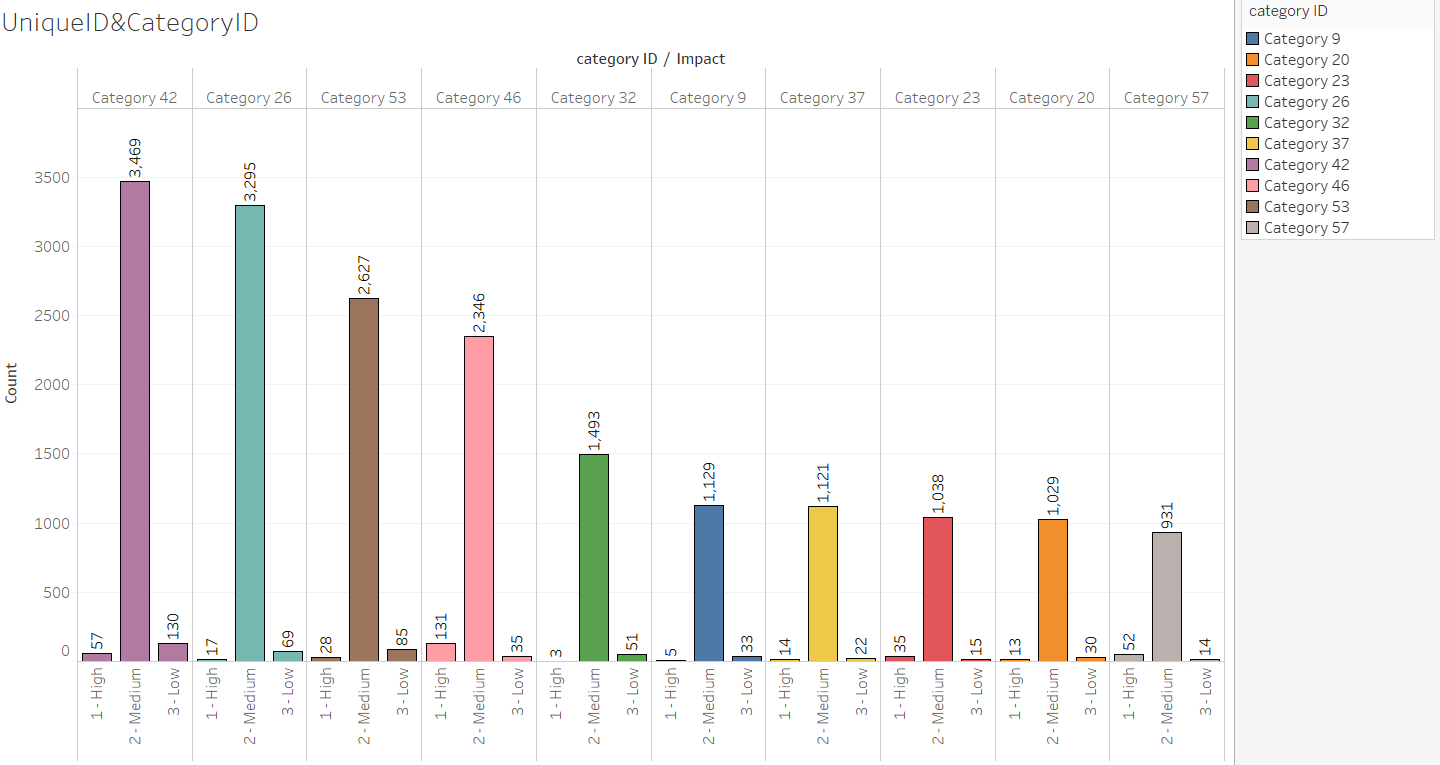
1. location -- There are 224 different locations



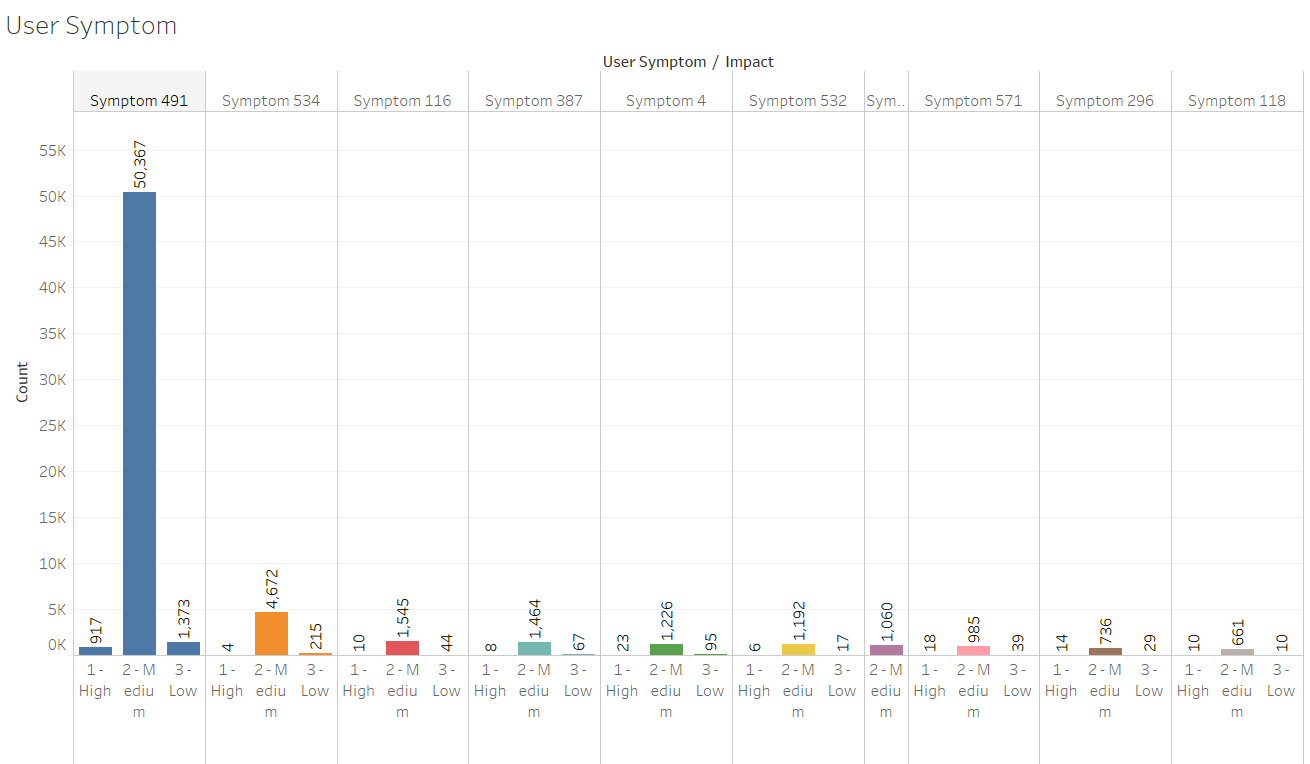


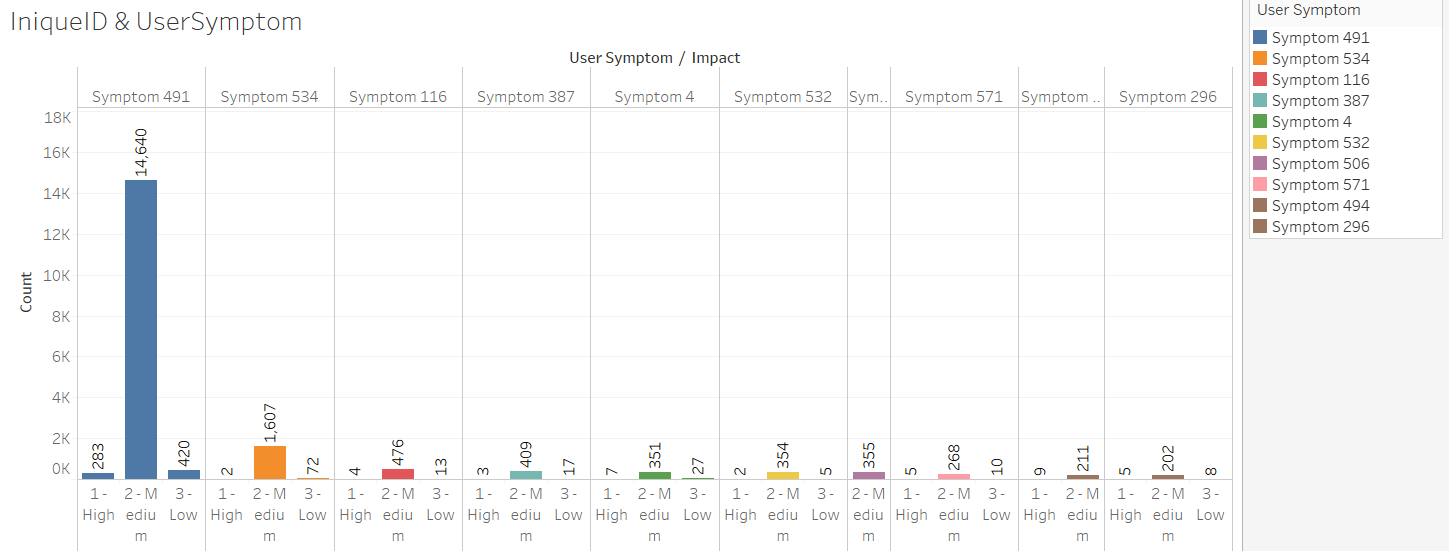
1. Category ID – There are 58 different category id



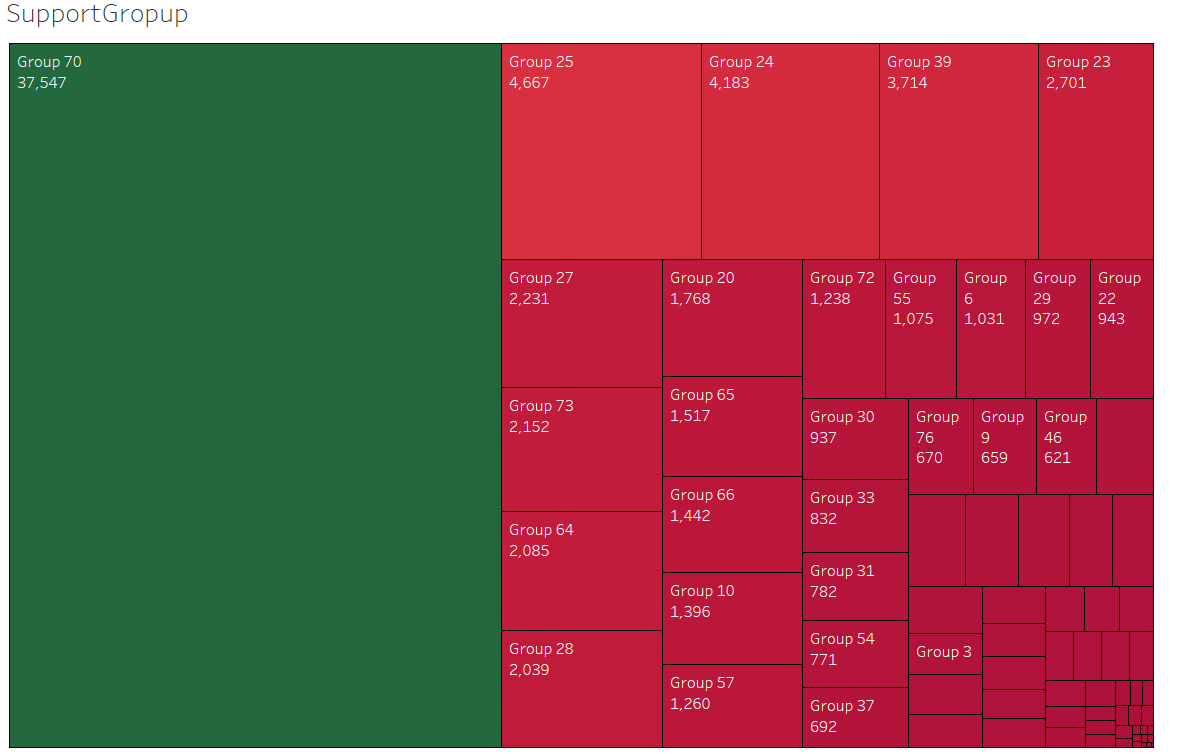


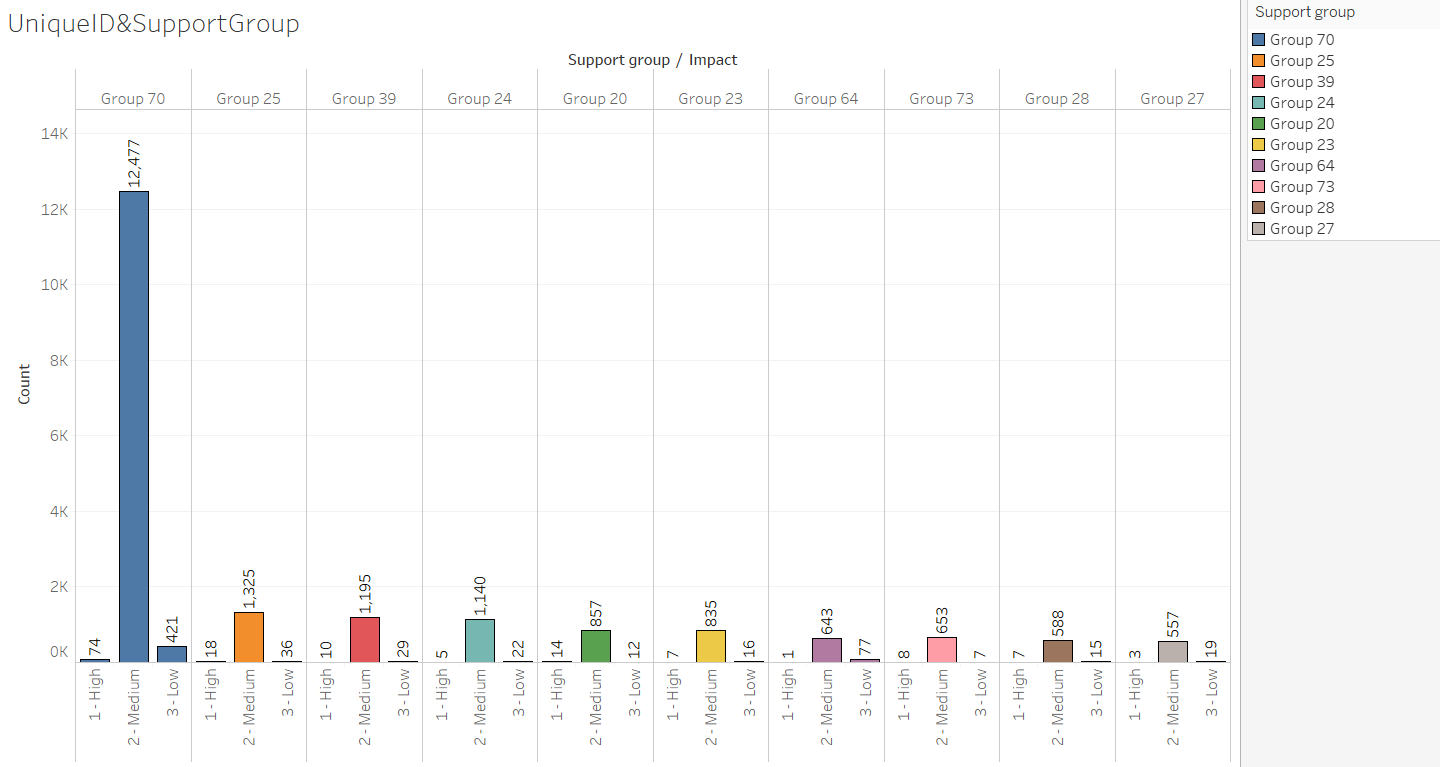
1. user\_symptom—There are 525 different user symptoms



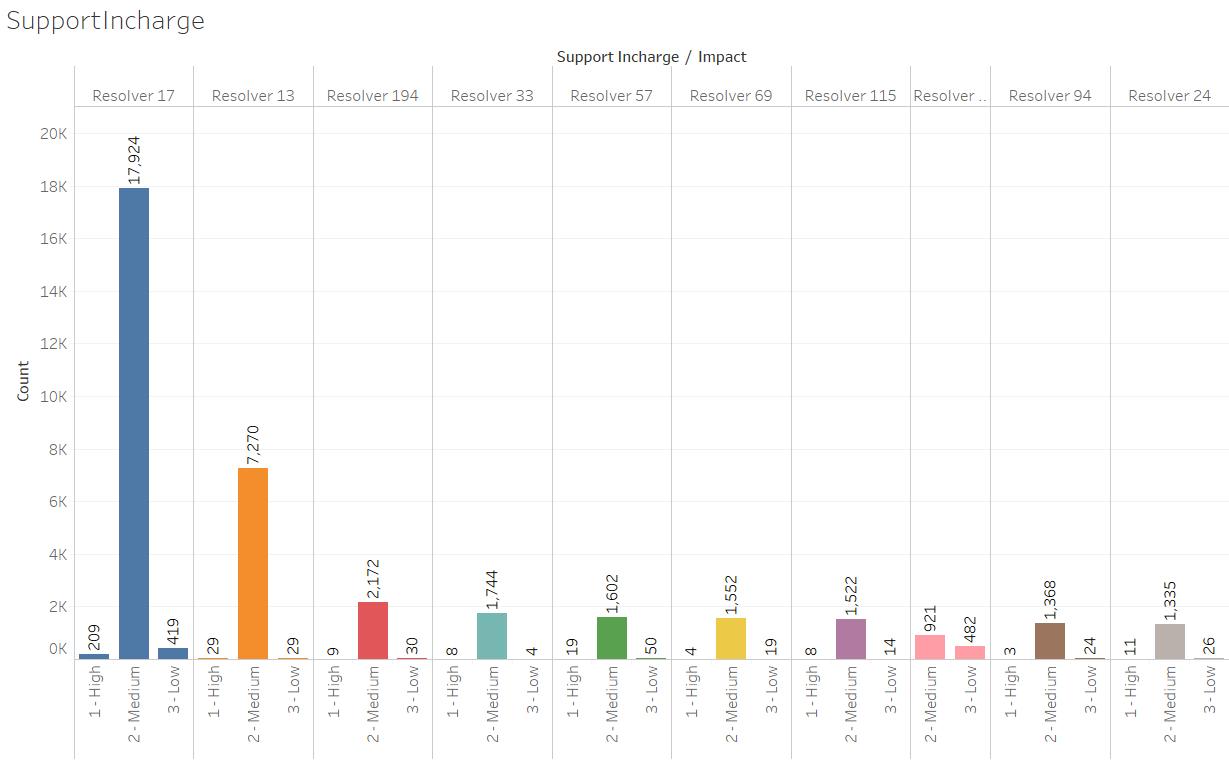


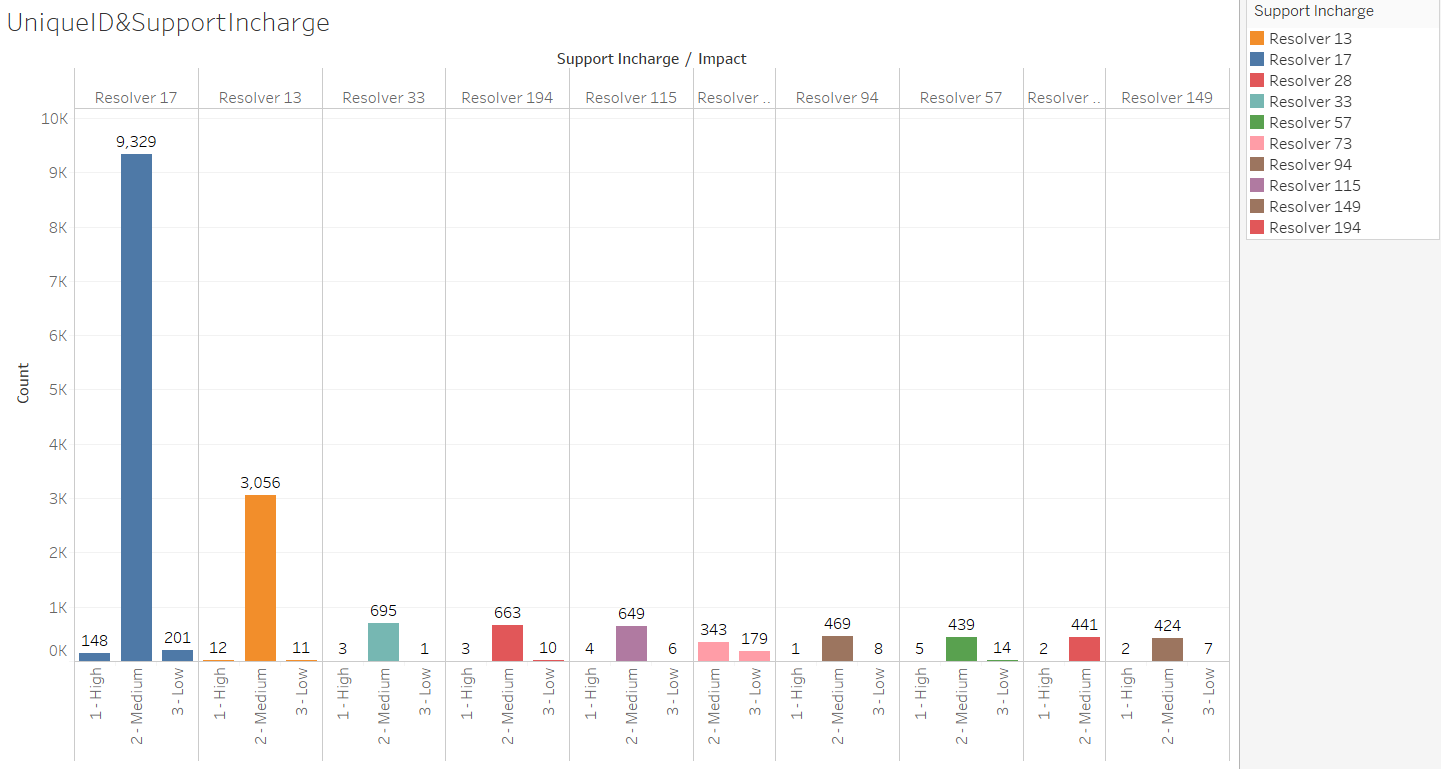
1. Support\_group There are 78 different support groups



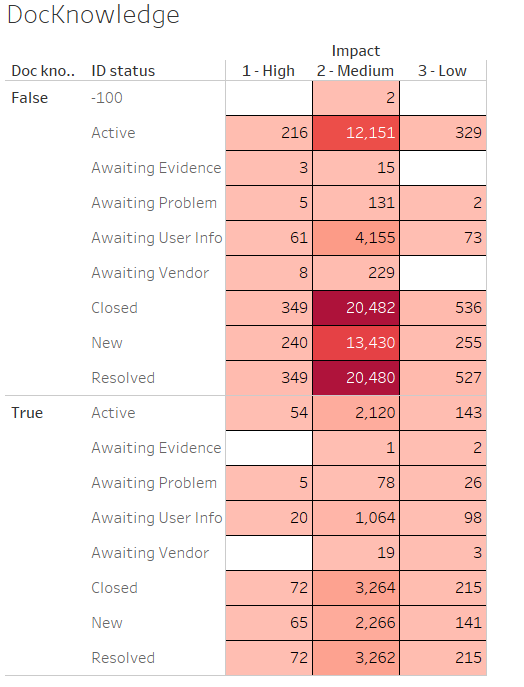
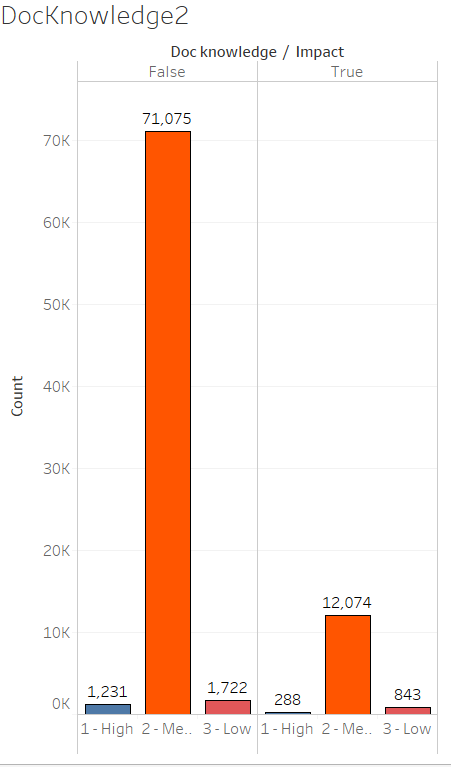


1. support\_incharge-- 234 unique support incharge

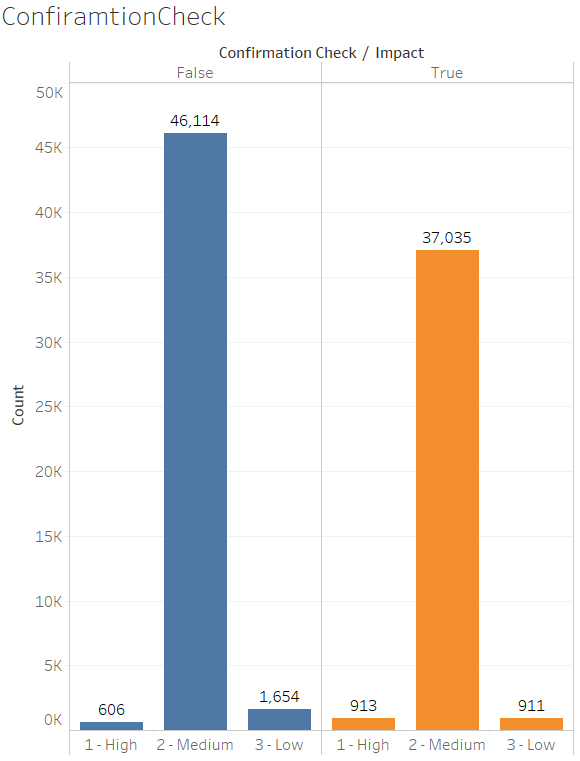
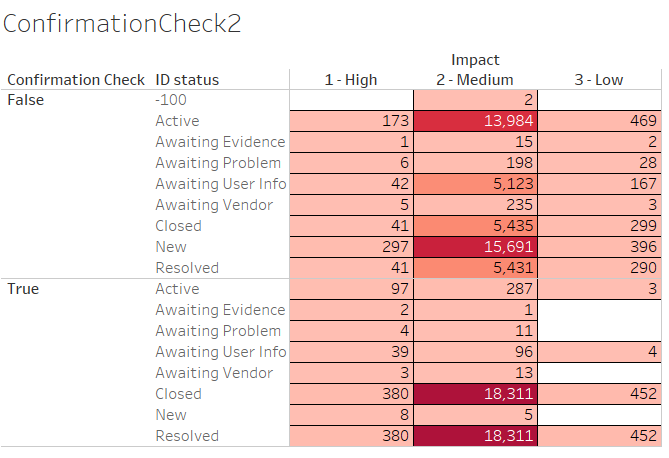




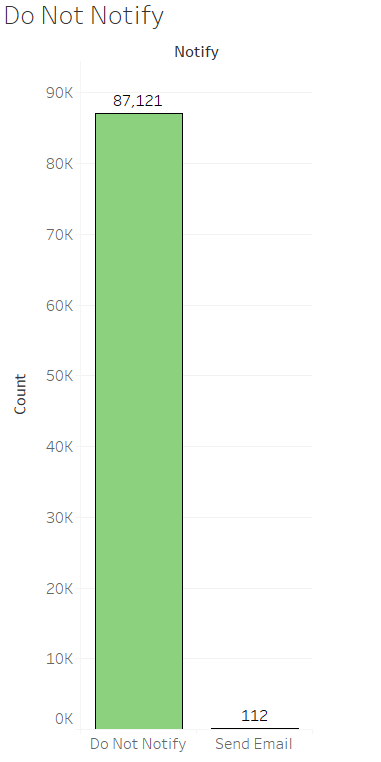
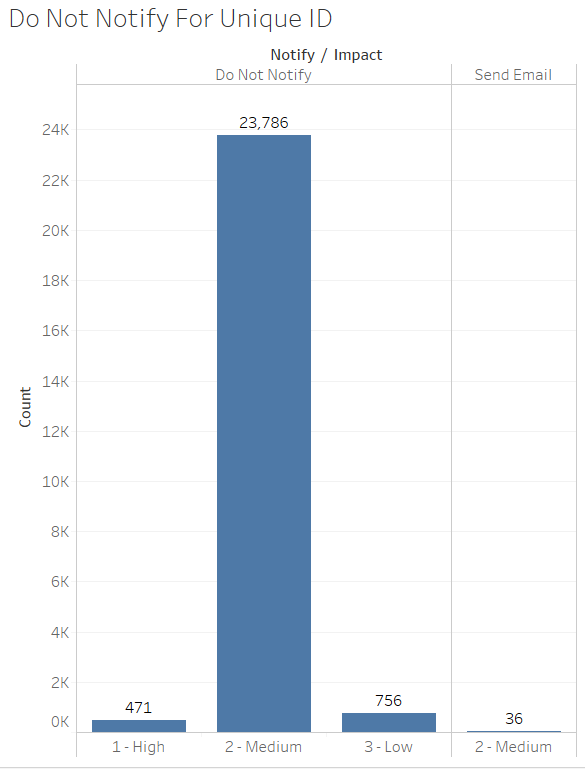
1. Doc\_knowledge ---- True, False

1. confirmation\_check True, False

1. notify Do Not Notify, Send Email

1. problem\_id There are 253 different problems



Around 85789 incident records belong to problem-ID Indicating by “?”. For which proper name is not mentioned. Have to deal with it separately.

1. change request 182, different change request



Around 86634 incident records belong to Change request indicating by “?”. For which proper name is not mentioned. Have to deal with it separately.

**Feature Selection**

Quantitative Variables like count\_reassign, count\_opening, count\_updated, opened\_time, created\_at, updated\_at do not have any influence on impact (Domain Knowledge). So, these variables are not taken into any consideration for further analysis.

**Check for association between Impact and other input variables**

**Chi Square Test**

H0: There are no relationship between two variables

H1 : There is significant relationship between variables of interest

After performing Chi Square Test we find the following results

**feature\_names P-Values Significant**

0 ID\_status 2.435171e-18 Yes

1 active 5.547200e-01 No

2 ID\_caller 0.000000e+00 Yes

3 opened\_by 0.000000e+00 Yes

4 Created\_by 0.000000e+00 Yes

5 updated\_by 0.000000e+00 Yes

6 type\_contact 2.068557e-133 Yes

7 location 0.000000e+00 Yes

8 category\_ID 0.000000e+00 Yes

9 user\_symptom 0.000000e+00 Yes

10 Support\_group 0.000000e+00 Yes

11 support\_incharge 0.000000e+00 Yes

12 Doc\_knowledge 4.354526e-146 Yes

13 confirmation check 1.952573e-51 Yes

14 notify 6.366832e-02 No

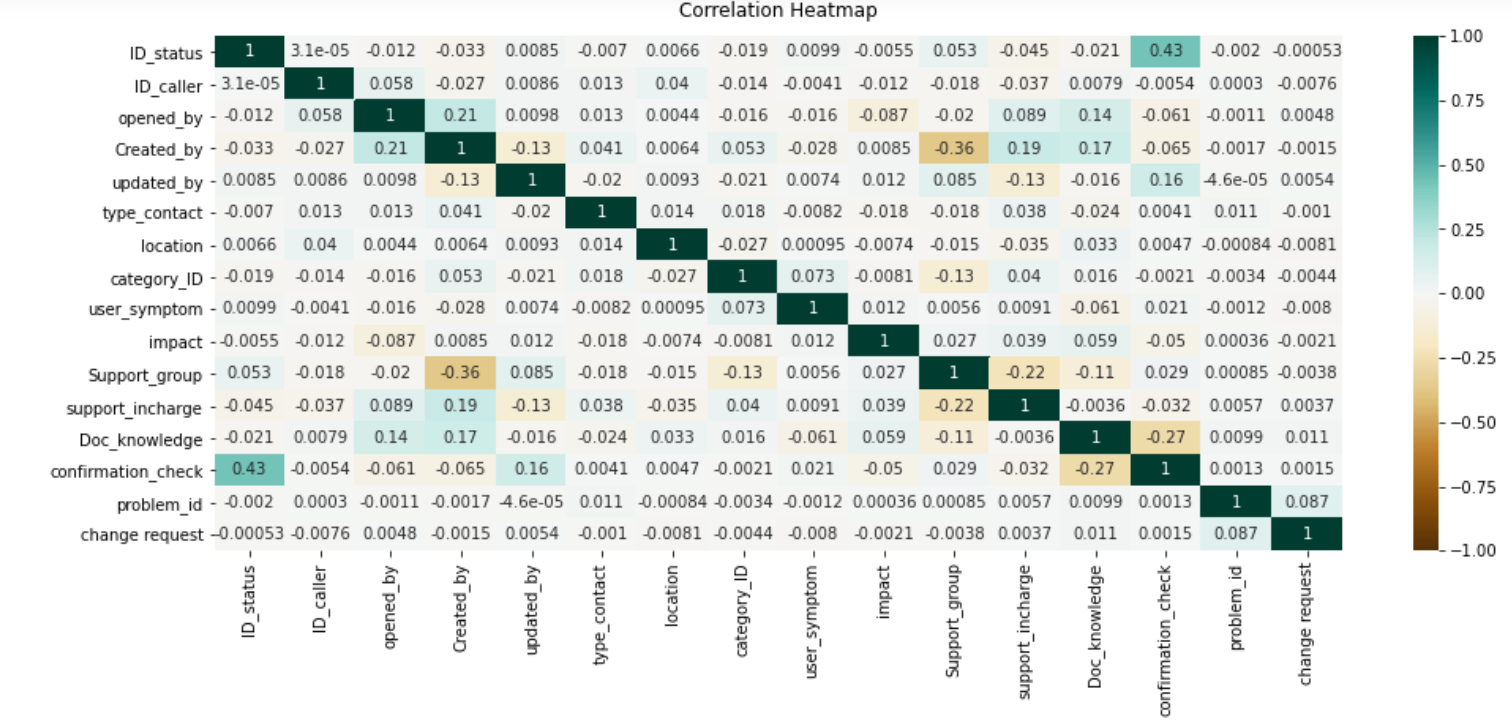
15 problem\_id 3.971051e-104 Yes

16 change request 1.169109e-96 Yes

From the p-values we calculated after Chi-Test , it is found that features like “active: and “notify” are not significant in this case and therefore not taken into consideration for further analysis.

Handling unwanted values and coulmns

1. Droping the Unwanted records from feature ID\_status where ID\_status =-100.



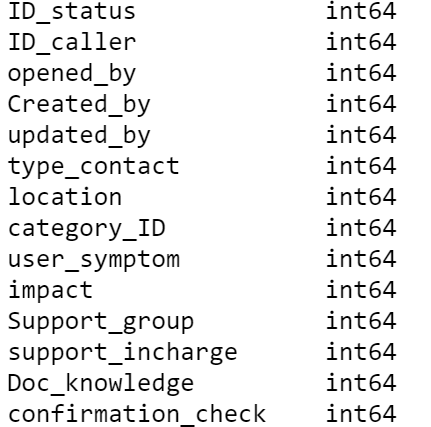
From the correlation heat map the following table can be summarized

|  |  |
| --- | --- |
| **Impact VS Input Variables** | **Correlation-Coefficients** |
| ID\_status | -0.0055 |
| ID\_caller | -0.012 |
| Opened\_by | -0.087 |
| Created\_by | -0.0085 |
| Updated\_by | 0.012 |
| Type\_contact | -0.018 |
| location | -0.0074 |
| Category\_id | -0.0081 |
| User\_symptom | 0.012 |
| Support\_group | 0.027 |
| Support\_Incharge | 0.039 |
| Doc\_knowledge | 0.059 |
| Confirmation\_check | -0.05 |
| Problem\_id | 0.00036 |
| Change\_request | -0.0021 |

From the above table, “problem\_id” and “change\_request” features have very low co-relation value. Moreover there are number of missing values is very high for these two variables. So these two variables are not considered for further consideration.

The following are the variables will be considered for further analysis

|  |  |
| --- | --- |
| **Input variables** | **Output Variables** |
| ID\_status | impact |
| ID\_caller |
| Opened\_by |
| Created\_by |
| Updated\_by |
| Type\_contact |
| location |
| Category\_ID |
| User\_symptom |
| Support\_group |
| Support\_incharge |
| Doc\_knowledge |
| Confirmation\_check |

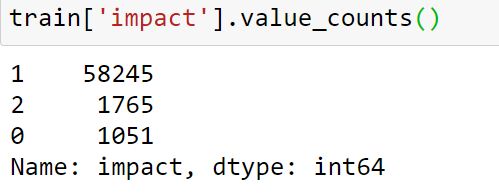


**Data Partitioning**

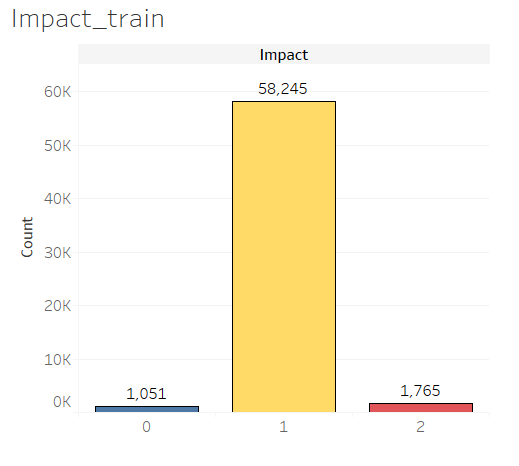
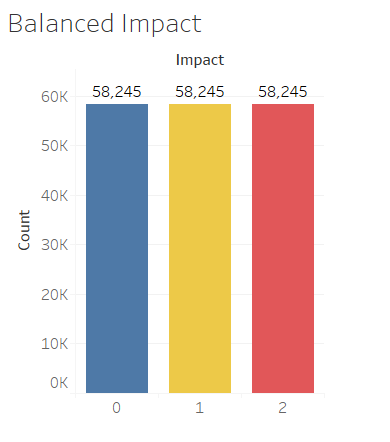
Dataset is now splited into train and test dataset by 7:3 ratio

**Balancing Imbalance datasets**

|  |  |
| --- | --- |
| **Categories** | **Value count** |
| medium = 1 | 58245 |
| low = 2 | 1765 |
| High = 0 | 1051 |



Since from the above snapshot, it can be found that there is no proper balance between different categories of target variable(impact) from train dataset, need to balance this imbalance in the dataset

SMOTE

**Model Building**