**Summary for Test Case 1**

1. Provide summary statistics of the data.
2. Syndicated requests for the week: 127647

Non-Syndicated requests for the week: 4473

Total Requests: 132120

1. Syndicated unique users: 127647

Syndicated requests per user: 1.103

1. Most requested RCID: 97, 12, 96, 1128, 40(mover, personal trainer, lawn care professional, house cleaner, carpet cleaner)

Least requester RCID: 815, 1012, 203, 202, 1179(Solar Panel cleaner, gazebo repairer , TOEFL tutor, series 7 tutor, math tutor)

1. Customers can submit their requests anytime of the week
2. All days of the week.

(the format in the csv file is mins:secs.ntn of a second, the whole dataset incurred around 160 hrs of requests right around 6.67 days)

1. Discern the trends in Data.
2. The major factor in for non-syndication is the Request status

|  |  |  |
| --- | --- | --- |
| Request Status | Non-syndicated Requests | Syndicated Requests |
| 1(Open) | 4393 | 101387 |
| 2(Closed/Cancelled) | 71 | 10833 |
| 3(Picked/Hired) | 9 | 15427 |

As you can see request status 1 has the highest non-syndicated request but it still has the highest Syndicated requests, with that in case request status 1 still has a good non-syndication standing due to the mere fact that it out syndicated other request status. Request status 3 has the lowest non-syndicated request and it has a relatively good value for syndicated requests. Request status 3 is the best for request status to for a request to be non-syndicated . At last Request status 2 has the 2nd highest non-syndicated requests(it leads request status 3 by a relatively significant margin 62) and the least Syndicated Request(lags status by a big margin 4593 ). It is the worst request status to have non-syndication.

1. Most non-syndicated RCID: 12, 1128, 96, 95, 97(Personal Trainer, House Cleaner, Lawncare Professional, Massage Therapist, Mover)

Least syndicated RCID: 605, 593, 565, 553, 1291(Meditator, Meditation Instructor, Jogging Instructor, Football Instructor, DJ)

If I had a little bit more time I probably have looked at the daily/hourly averages of certain parameters such as requests, syndicated requests, non-syndicated requests. The most and the least syndicated jobs by ratio(# of requests with that RCID /# of syndicated request with that RCID), overall syndication ratio. for with these values you will be able to know which jobs are in demand-has a good conversion rates(request syndicated) that will be valuable in the business. I would just like to point out that it is better if the timestamps used in this dataset would be of standard format.(I did some extra work it shows the jobs that have a good and bad syndication rate (in the AUX part of my python program))

|  |  |  |  |
| --- | --- | --- | --- |
| Good | Syndication Rate(G) | Bad | Syndication Rate(B) |
| 1192 | 1 | 132 | 0.65 |
| 1187 | 0.9958 | 960 | 0.64 |
| 180 | 0.9952 | 350 | 0.6 |
| 1200 | 0.9935 | 676 | 0.5 |
| 1191 | 0.9935 | 150 | 0.0 |

Range: 0 <= Syndication rate <= 1

**Summary for Test Case 2**

1. Provide summary statistics of the data.

Cells with null values: 3620

CASE:

No. of cases: 6177

No. of unique cases(null included): 6176

No. of unique cases(nulls excluded): 6175

CASE ORIGIN:

We have 8 categories of Case Origin.

Live Ops Inbound Support Call, Live Chat, and Text Message(2417, 1427, 1019) are the main means of inquiry, and makes up 78.73% of all queries.

What’s interesting is that web-based queries make the most portion of all queries(Live Chat, Email, Web) excluding social media, with a total of 2713 queries, making up 43.92% of all queries.

Average query/Case Origin: 772.125

Standard Deviation: 846.32

CASE CATEGORY

We have 21 categories for Case Category.

Requests, Account Profile, and Refunds are the most handled cases(1180, 1152, 1050). It is interesting to point out that Requests, Accounts / Profile, Refunds, Quotes, Reviews, and Credits/Pricing/Payments makes up a 5328 cases equivalent to 86.26% of all cases, a huge portion.

Average query/Case Category : 294.14

Standard Deviation: 418.23

CSAT OWNER: NAME:

We have 228 agents handling 6177 queries.

Travis Oliver, Devin Lane, and Sarah Levorsen are the agents with the most queries(107, 103, 94)

Average query/agent: 27.09

Standard deviation: 24.58

TEAM:

We have a total of 27 teams.

TPH Chat, TPH Tier 4, adnd TPH Text are the teams with the most queries(1253, 1126, 796)

Average query/team: 228.78

Standard Deviation: 324.03

REPORTING MANAGERS:

We have 30 Reporting Managers

Average query/manager: 205.9

Standard deviation: 317.18119843975325

OFFER DATE AND OFFER RESPONSE DATE:

Queries are done in a period of: 64 days

Queries are responded in a period of: 63 days

Average time till an action from our agents takes place: 1311.74 mins(Offer response date – Offer Date)

Standard Deviation: 17876.51073764476

NUMBER OF INTERACTIONS TO RESOLUTION:

|  |  |
| --- | --- |
| Once | 1756 |
| More than once | 512 |
| My issue is not resolved | 343 |
| 1 | 2 |
| 4+ | 2 |

CSAT Response:

subpar = CSAT ratings from 1 - 3

percentage bad/subpar responses: 20.96

percentage good responses: 79.04

Average CSAT rating: 4.24801683665 / 4.25

Due to time constraints I chose to analyze the effects of 5 attributes and its effects on the CSAT rating. These attributes are Case Category, Case Origins, Action Time(The time between Offer Response Date and Offer Date), Teams, and Number of Interactions to resolution. And I put a premium on the subpar ratings(CSAT ratings 1- 3)and how these attributes contribute to the low CSAT ratings. Let’s first talk about Action Time(as seen in the “avg\_response\_time.csv”)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| Average response time in(mins) | 4688.677 | 2133.427 | 2876.821 | 871.1125 | 721.5613 |

According to the data subpar(CSAT ratings 1 - 3) have higher average response time than that of good ratings(CSAT ratings 4-5). It means that it took more time for our agents to respond to query. Especially in the case of CSAT rating 1 that has an average response time of 4688.677 mins, this service that is graded the worst takes 3892.34 minutes longer than good services (796.34 mins on avg) to be acted upon. It seems that the time between a customer asks for support and the time our agents act on it somehow affects how the help will be graded.

In the Case Category I only considered scores of queries that to fall into these 4 categories: Account / Profile, Quotes, Refunds, & Request as seen in “case\_categories.csv” and “case\_categories1.csv”. Since most queries fall into these 4 categories (69.26%) and the 17 categories make up the rest. The percentage of the average subpar categories that fall into these categories is 17.11% slightly lower than the overall percentage bad/subpar responses which is 20.96%, that means these categories most likely do not increase the subparity of total queries and therefore Case Category has a low effect on the subparity of the total queries. However there are categories that have a higher subparities. Queries under Credits/ Pricing / Payment and Reviews are quite sizeable and they have bit higher subparity percentage(around 25%). Queries under Marketplace Integrity, Other, and everything under ProAssist have few queries-their subparity percentage is high 58.33%, 35.24%, 39.17%(average on this one) meaning to say that most queries fall into both of these are most likely given poor customer service. Even though in general Case Category have little effects on the subparity, there are some categories that shows to have given subpar customer service.

In the Case Origin attribute I only considered scores of queries that fall in these categories: Live Chat, Live Ops Inbound Support Call & Text Message as seen in "case\_origin.csv" and "case\_origin1.csv". Since most queries fall into these 3 categories(78.73%) and the remaining 5 make up the rest. The percentage of average subpar categories that fall in these 3 categories is 17.26% slightly lower than the overall percentage bad/subpar responses which is 20.96%, that means these categories most likely do not increase the subparity of total queries and therefore Case Origin has a low effect on the subparity of the total queries. But there are categories under Case Origin that that I’d like to point out because of their higher subparity percentage. I singled-out web based queries Email, Web and Social Media(excluding Live Chat) due to the fact that people now commonly use the internet for most of their transactions. These categories have poor CSAT scores (35.35%, 34.84%, 50% subpar percentages respectively)as of now these high numbers overall might not yet affect our overall subparity percentage, but in the future as more people use the internet as its medium for customer support; if these numbers stay the same our CSAT scores might go lower and customer service may seem to have suffered.

Let’s now look at Team, an attribute that is related to workforce performance (refer "team.csv" and “team1.csv”). Overall our Teams’ performance is actually almost the same as our overall subparity percentage (20.92% as 20.96%). That mean our Teams’ performace is strongly translates to the quality of our customer service. But I digress, now we ask the question does the Team that handles your query affects the outcome of you query at the same time it also the each individual team’s performance. Let’s first look at team TPH and break it down. TPH has 5 teams under it and handles 3206 queries overall, TPH team is a team that handles most of our queries that’s why it’s critical for us to look at it. Now let’s breakdown each team’s performance. TPH Chat and TPH Text handled a good portion of the queries handled by TPH and they got good subparity percentage(15.48 % and 13.19%). TPH tier 4 handled the 2nd most queries of all the TPH teams and got a slightly bad subparity percentage(27.09%). TPH teams has an average subparity percentage rating(WA) 19.62% meaning to say this team still performs within our standards. Next we take a look at Support Team. This team handles 2378 queries and has an average subparity percentage of 13.1%-relatively lower than our overall subparity percentage, that means this team is performing good customer service. Lastly let’s look at the worst perfoming team, Analyst. They have 98.17% subparity percentage, meaning to say almost all queries that they’ve handled were not resolved. Like in the Case Origin, Social Media got a bad subparity percentange(47.83%). Social Media seems to be a weak point of our customer service.

Now to the last attribute that we will analyze is the Number or interactions to resolution.

|  |  |  |
| --- | --- | --- |
|  | subpar\_pct | No\_of\_interactions\_to\_r\_val\_counts |
| 1 | 50 | 2 |
| 4+ | 100 | 2 |
| More than once | 31.445312 | 512 |
| My issue is not resolved | 70.553936 | 343 |
| Once | 6.378132 | 1756 |

By looking at the table above it is tempting to say that the lower the number of interactions the low CSAT scores will be given less frequently and the service will be graded . Once has a subpar percentage of 6.34% whilst More than once and 4+ has subpar percentages 31.45% and 100% which clearly shows that the lower the interactions the better the. But let’s look at the bigger picture here Number of interactions to resolution have 3562 empty cells w/c makes up mostly of the missing values in our dataset. Meaning to say that queries that don’t have values in their Number of interactions to resolution and w/ their corresponding CSAT scores are under-represented.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| Queries with Graded CSAT scores w/ corresponding null values in their Number of interactions to resolution column | 325 | 227 | 225 | 335 | 2450 |
| Queries with Graded CSAT scores | 573 | 358 | 364 | 551 | 4431 |
| Percentage of Graded CSAT queries w/ empty values | 56.72% | 63.41% | 61.81% | 60.80% | 55.29% |

As you can see in the table above most CSAT graded queries have no values for their Number of interactions, the under representation on each CSAT graded query is above 50%, making those present values for Number of interactions non-representative. The under representation made me decide not to make an analysis based on “Number of interactions to resolution”even if there are values that are present.

**Conclusion & Recommendations:**

Given the time constraint and the amount of options you can choose to deconstruct the data, I’ve limited my analysis to a few attributes and just made a straight forward analysis on how they affect the CSAT scores given to each query. But If I am more oriented on the given data I would have an in-depth analysis of each attribute and probably used more weighted averages. I also plan to make Linear / Logistics Regression models in R, but since I did my analysis in Python(pandas) it isn’t possible at the moment. The most certain driver of CSAT score that I’ve found is the “Action Time”. There seems to be an established inverse relationship between the time that a query has been made to when it was acted upon and the CSAT score(longer waiting time till our agents take action to a query). Whilst other attributes such as Case Origin, Case Category, Team have little effects on CSAT scores. The short term solution that I can recommend is for Thumbtack to shorten the time between query request and query action. In my limited analysis I’ve pointed out specific categories with high subparity percentage wherein we can take a look and make some adjustments. But moving forward in the future I have to make a premium on some metrics that I believe will be the crucial in the quality of our customer service. First is workforce performance. Given this data set workforce performance can be divided into per team and per agent. Let’s first talk about per team performance. I’ve already had a breakdown on Team’s performance above now I want to talk about the 2 teams that currently handles most of the customer queries Support Teams and TPH. Support teams are doing great having a very low subpar scores for their queries. TPH on the other hand is special since it handles most of the queries. We should look at the individual teams under TPH and see where there is a problem. Like TPH Tier4 wherein 27% of the time their handling of queries is subpar. This is the type of scrutiny going forward that we must to on all teams to make the customer service better. Next is individual performance. We should scrutinize each agents performance wherein we will look on their individual CSAT scores and possibly other metrics to asses them accordingly, if we see lapses in their performance and it translates to poor customer service we must take proper actions. Going forward I believe that work force performance is one of the key drivers for customer service satisfaction. The next driver I want to point out is the medium wherein the query was made, case origins. In the future a bulk of our queries will be through web based mediums Live Chats, Emails, and Social Media. Right now thumbtack web based services have poor customer service ratings except for live chat that’s performing a little bit better than our overall performance. Thumbtack should now look on how queries through this medium are being dealt. Email and web-based customer service mediums have a 35% subparity percentage and Social media has a 50% subparity percentage-meaning to say half of social media queries are bad customer service. Thumbtack should be aware of this pain points since thumbtack is an internet based service and that these drivers are important not only for customer service but as well to the company’s overall performance. These are the trends that I’ve seen and I’ve given some recommendations where there seems to be lapses. Thank you for considering my application I hope we can work together soon!