

# Telecom: Identifying Ineffective Operators

By: Aggie Benichou  
Yandex100 Data Analyst





## Task

A virtual telephony service is developing a new function that will give supervisors information on the least effective operators. An operator is considered ineffective if they have a large number of missed incoming calls (internal and external) and a long waiting time for incoming calls. Moreover, if an operator is supposed to make outgoing calls, a small number of them is also a sign of ineffectiveness.

Identify ineffective operators based on the number of calls (incoming / outgoing), the waiting time for incoming calls and the number of outgoing calls (if the operator is supposed to make them).



# Data

There were two datasets provided:

- On the left, a dataset containing call information such as: the user ID that called, the date they called, the operator ID that the call was forwarded to, the call duration, the wait time, etc.
- On the right, a dataset containing user information such as: the user ID, the tariff plan they are subscribed to and the date of registration.

|       | user_id | date       | direction | internal | operator_id | is_missed_call | calls_count | call_duration_sec | total_call_duration_sec |
|-------|---------|------------|-----------|----------|-------------|----------------|-------------|-------------------|-------------------------|
| 1     | 166377  | 2019-08-05 | out       | True     | 880022      | True           | 3           | 0                 | 5                       |
| 2     | 166377  | 2019-08-05 | out       | True     | 880020      | True           | 1           | 0                 | 1                       |
| 3     | 166377  | 2019-08-05 | out       | True     | 880020      | False          | 1           | 10                | 18                      |
| 4     | 166377  | 2019-08-05 | out       | False    | 880022      | True           | 3           | 0                 | 25                      |
| 5     | 166377  | 2019-08-05 | out       | False    | 880020      | False          | 2           | 3                 | 29                      |
| ...   | ...     | ...        | ...       | ...      | ...         | ...            | ...         | ...               | ...                     |
| 53895 | 168606  | 2019-11-09 | out       | False    | 957922      | False          | 4           | 551               | 593                     |
| 53896 | 168606  | 2019-11-10 | out       | True     | 957922      | False          | 1           | 0                 | 25                      |
| 53897 | 168606  | 2019-11-10 | out       | True     | 957922      | True           | 1           | 0                 | 38                      |
| 53898 | 168606  | 2019-11-11 | out       | True     | 957922      | False          | 2           | 479               | 501                     |
| 53899 | 168606  | 2019-11-15 | out       | True     | 957922      | False          | 4           | 3130              | 3190                    |

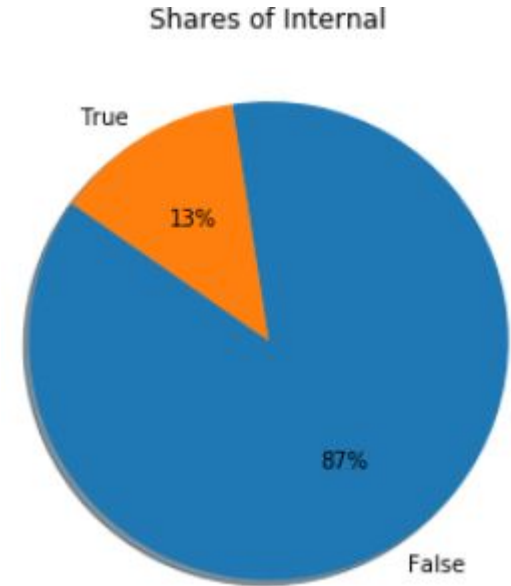
41491 rows × 9 columns

|     | user_id | tariff_plan | date_start |
|-----|---------|-------------|------------|
| 0   | 166713  | A           | 2019-08-15 |
| 1   | 166901  | A           | 2019-08-23 |
| 2   | 168527  | A           | 2019-10-29 |
| 3   | 167097  | A           | 2019-09-01 |
| 4   | 168193  | A           | 2019-10-16 |
| ... | ...     | ...         | ...        |
| 727 | 166554  | B           | 2019-08-08 |
| 728 | 166911  | B           | 2019-08-23 |
| 729 | 167012  | B           | 2019-08-28 |
| 730 | 166867  | B           | 2019-08-22 |
| 731 | 166565  | B           | 2019-08-08 |

732 rows × 3 columns

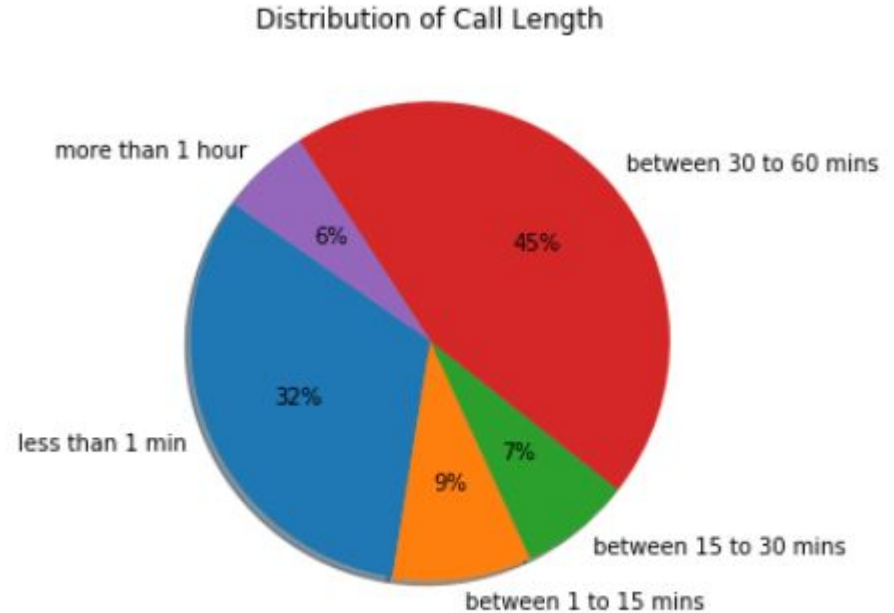
# Shares of internal and external calls

- The adjacent pie chart shows the shares of calls found in the dataset, which contains internal calls (within the company) and external calls (from outside the company, such as from users).
- According to the data, about 87% of calls found within the dataset are external calls.
- Most calls are from users to the virtual telephony service



# Distribution of call length

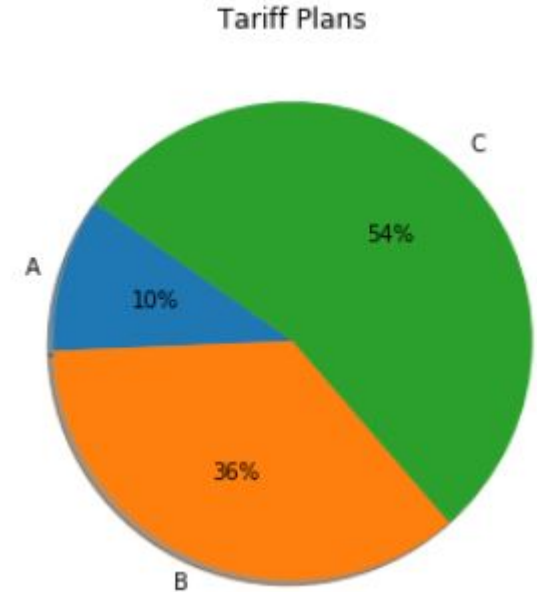
- The adjacent pie chart shows the distribution of call lengths for calls within the dataset.
- Each call was classified within one of five categories: less than 1 minute, between 1 to 15 minutes, between 15 to 30 minutes, between 30 to 60 minutes and more than 1 hour.
- According to the data, almost half of the calls are between 30 to 60 minutes but the next largest category are calls that lasted less than 1 minute.





# Tariff Plans

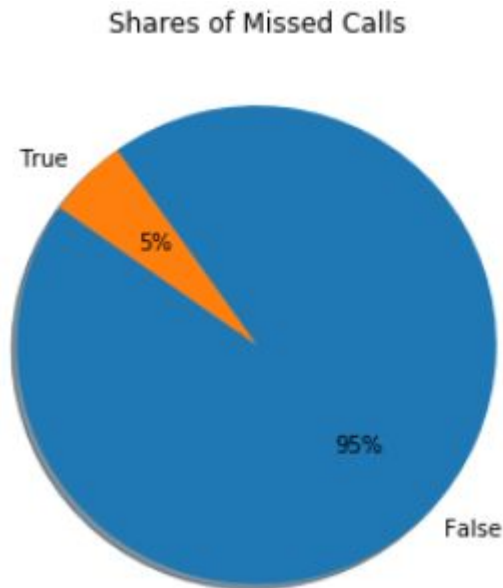
- The adjacent pie chart shows the distribution of tariff plans that users are subscribed to within the dataset.
- Out of the three available plans, more than half of the users are subscribed to tariff C.
- The next most popular plan is tariff B and the least popular plan is tariff A.
- Details on these tariffs are not provided.





## Shares of missed calls

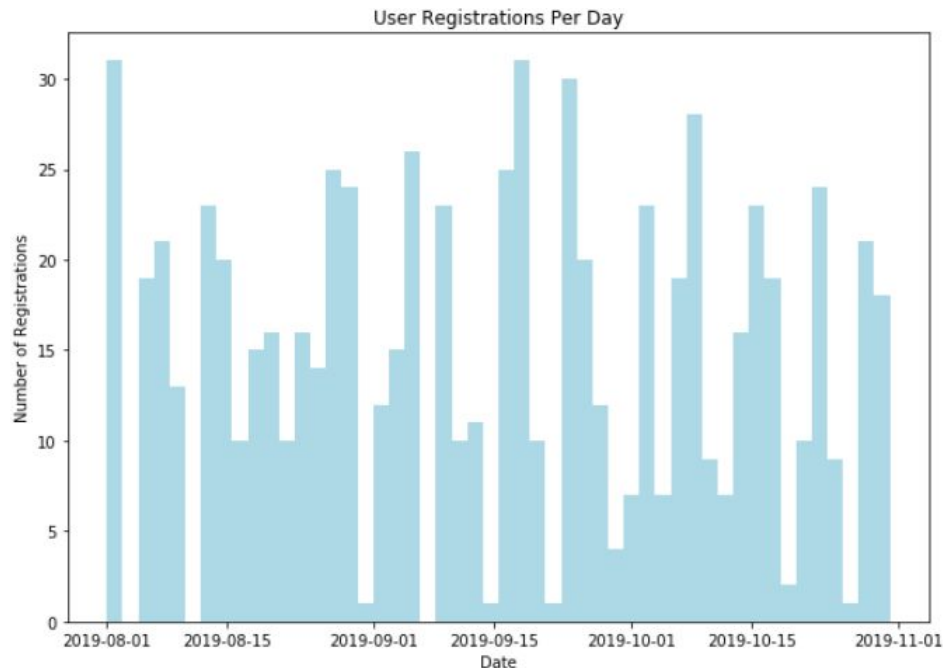
- The adjacent pie chart shows the share of missed calls in comparison to the total number of calls.
- According to the data, almost 95% of calls are not missed calls. This is a positive feature.





# User registrations per day

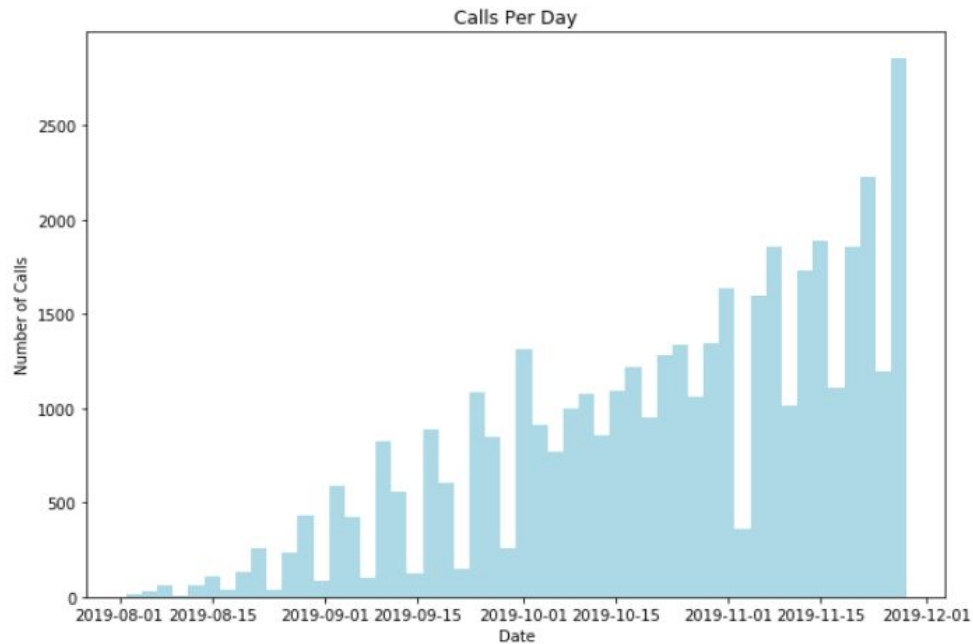
- The histogram shows the distribution of user registrations to the service per day within a 3 month period.
- The graph shows no pattern or trend to user registration: some days have more than 30 registrations while other days have less than 5 registrations.
- There is also no pattern as to the time of month: sometimes the beginning of the month have high registration, sometimes the end of the month.





# Number of calls per day

- The histogram shows the distribution of number of calls (internal and external) per day within a 3 month period.
- The graph shows an upward trend in the number of calls which could be because user registration increased over the same time period, thus increasing phone calls in order to keep up with demand.
- Another interesting point to notice about the graph is that there are dates where the number of calls is quite high and then the day or few days after, it drops.





## Identify operators with a long wait time.

- The average wait time that a user waits per call is 97 seconds but to identify operators with a long wait time, I pinpointed operators within one standard deviation of the average.
  - This covers over 68% of all the data points.
- I grouped individual operators with their average wait time and compared the total average wait time to one standard deviation limits.
  - The operators that were outside of the one standard deviation from the average were identified as ineffective.
- This method identified 22 operators with a long average wait time, their wait times ranged from 5 minutes to over 15 minutes.



# Identify operators with many missed calls.

- The average number of missed calls per operators is 3 missed calls but to identify operators with a long wait time, I pinpointed operators within one standard deviation of the average.
  - This covers over 68% of all the data points.
- I grouped individual operators with their average number of missed calls and compared the group average number of missed calls to one standard deviation limits.
  - The operators that were outside of the one standard deviation from the average were identified as ineffective.
- This method identified 24 operators with many missed calls, their number ranged from 7 to 28 missed calls per operator.



# Identify operators with few outgoing calls.

- The average number of outgoing calls per operator (who is supposed to be making outgoing calls) is 36 calls but to identify operators with few outgoing calls, I pinpointed operators within one standard deviation of the average.
  - This covers over 68% of all the data points.
- I grouped individual operators with their average number of outgoing calls and compared the group average number of outgoing calls to one standard deviation limits.
  - The operators that were outside of the one standard deviation from the average were identified as ineffective.
- This method identified almost 500 ineffective operators with few outgoing calls.



## Determine the number of ineffective operators.

- According to the company's definition, an ineffective operator is one who:
  - Has a long wait time AND many missed calls (for incoming calls).
  - Has few outgoing calls (for outgoing calls).
- Out of the 1092 total operators, this logic identified a total of 143 ineffective operators as defined by the company's measures:
  - 4 operators ineffective for incoming calls.
  - 139 operators ineffective for outgoing calls.



# Hypothesis: Average call duration for ineffective operators and effective operators are the same.

- To test the hypothesis that the average call duration for ineffective operators and effective operators are the same, an independent samples t-test to compare the means from the two groups (ineffective operator average call duration and effective operator average call duration) can be used.
- The result of the t-test suggests that we should reject the hypothesis, thus indicating that the average call duration of ineffective operators and effective operators do differ enough to be statistically significant.
- This can be further proven by looking at the numbers: the average call duration of ineffective operators is 905 seconds (15 mins) and the average call duration of effective operators is 648 seconds (11 mins).



# Conclusion

- To summarize, the analysis identified 143 ineffective operators (according to company measures):
  - 4 operators ineffective for incoming calls.
  - 139 operators ineffective for outgoing calls.
- Action against the 4 ineffective operators for incoming calls should be taken immediately, as these operators have had a long average wait time and many missed calls over a 3 month period.
- However, further analysis should be done for the 143 ineffective operators for outgoing calls. I do not believe that all of these operators are performing badly.