Exploratory data analysis

Contents

Challenge and variables description

Problem

The number of agents available is crucial to the Contact Center success operation, where a low number of agents will lead to long waiting queues for the contacts and a high number will increase the operation cost by having agents doing nothing.

The dataset corresponds to Contact Center metrics aggregated in 15 minutes intervals for a real Talkdesk client between 2017-01-01 and 2018-11-30 (23 months), and the competition goal is to forecast the number of agents for the days between 2018-12-01 and 2018-12-31 (1 month).

Features

The following information is provided with the data:

- *Interval* - the date and time corresponding to the contact center activity. This is given on a 15 min interval.

Events:

During the initial presentation we learnt that talkdesk runs an "event" based operation. **A call is an event, and event triggers actions. The duration of each event is a restriction to the number of workers.**

For each event information about quantity and duration is given:

- *Total Calls* - the number of calls. This represents the total amount of calls received during the time period. For example between 2017-01-01 00:00:00 and 2017-01-01 00:00:00 this specific costumer received 227 calls.

- *Total Calls Duration* - the duration of a calls in seconds. This means that on the 2017-01-01 00:00:00 a total of 227 where received and took the total of 11880 seconds or in hours 3.3

- *Abandoned Calls / missing_calls* - the number of calls that reach Talkdesk platform but hung up before it was answered by an agent. No further information regarding the reason why the call was missed is given. This may or may not impact the solution. For example, if miss calls are due to all operators being occupied, then, in order to maximize the service it may pay off the hire an extra worker.

  This feature raises the question of how long is an event. It makes sense that we take into account that one event requires a 1. setup time, an 2. execution time, a 3. after event task and probably a 4. cool down between events.

Workers:

Being a service, "events" take time as raw material to be completed. This time is provided by workers which can supply a limited amount of it per day. The data provides us information regarding quantity for each time interval (headcount and available time) and how it is distributed by "tasks" or "status".

From the information provided we can deduce that a worker can be at the following "status" at a given time, 1. Available, 2. Away, 3. Busy, 4. On a call, 5. After Call Work (ACW). Without further information we assume this are the only possible status and therefore the sum of all 5 should result on the total available time at a given time period.

- *Available Time* - agents time *online* and free to take a new call. During this time a task/event was not assigned to an agent. This is lazy time, therefore the smaller the better.

- *Away Time* - agents time *online* and away from the computer unable to receive a new call. It includes non productive chargable time. This time the agent is "active" but "unavailble". He/she could be away at the bathroom, having a coffee breake etc. Althought the minimum the better we consider this time to be a necessity and therefore given data to be taken into consideration.

- *Busy Time* - agents time online and busy with tasks or meetings not related to handling calls. That means admin time and non billable.

- *On a Call Time* - agents time online and handling (talking) with a contact

- *After Call Work Time* - agents time online and closing the contact interaction, for example, update ticket status on CRM systems

- *Total Handle Time* - total agents time from the point a call is answered until after the agent closes the contact interaction. This is the sum of "On a call Time" and "After Call work time"

  Total billable time = Total Handle Time Total non billable time = Available Time + Away Time + Busy Time Where the non billable is a consequence of either lack of events (Available time), admin tasks (busy time) or temporary unavailability

Agent Headcount - total staff count for the interval (target). **side note: given the information this might represent the average head count**

Metrics:

The data provided already includs some metrics used by talkdesk

- *Occupancy Rate* - agents time in an available status or handling a contact divided by the total agents time. we can extract the total agent time from this ratio.

- *Utilization Rate* - agents time actually handling interactions divided by the total agents time
- *Shrinkage Rate* - agents time occupied with not-handling interactions related tasks divided by the total agents time. This means *Busy and away time divided by available time.* The available but not assigned is not included.

link to competition

Exploring data

Data consistency

With the above feature definition lets make some consistency checks, mainly focusing on the amount of time available and used. For this a new table was created with the following values:

- *numberAgents_NoCalls*: given the total calls duration in seconds, if all available time was call time, how many agents would be needed. This result does not take into consideration the after call work time (ACW) and is calculating by diving the total_calls_duration by 900 seconds (the number of seconds by each period).

- *agentsAvailableTime*: taking into account the headcount at any given time period

Single features distributions

Converting to time series

```
# convert to time base

timeSeries_data <- xts(x = data, order.by = as.POSIXct(strptime(data$interval,"%Y-%m-%d %H:%M:%S"), tz =
timeSeries_data$interval <- NULL
```

QUESTIONS:

- How does the headcount turnover works? for example, if we have a pick at mid-day but no calls during the rest of the day is it required that we keep all individuals during that day?
- Total calls are only calls answered? Are abandoned calls already included as total calls?
- What is the difference between on_a_call_time and total call durations? Given that the first surpasses the first does it includes the call setup? [providing that the time count starts automatically]
- How is it possible that the available time for each time signature is smaller than the active time?
- Does the agent headcount feature represent the average headcount?