**Home Instead IT Data Analyst Assessment Report**

Presented by*:* **Balaganesh Chakrahari**

Coordinator*:* **Melissa Thorne**

Git*:* <https://github.com/balaganesh99/Home_instead>

**Inroduction**

The fictitious company ACME Corp is interested in understanding more about their business, and they have enlisted your services as a data analyst to help. I will be using a compressed CSV file that represents fictitious historical data from ACME Corp relating to appointments between its customers and service providers.

Programming Language used – Python (Pandas , matplotlib , NumPy , Seaborn)

Dataset: <https://goo.gl/0vCT8p>

The data set is tuned per the required analysis with the following attributes.

Attributes

* ‘ID’ – The unique identifier for this appointment
* 'Start’– The UTC timestamp when the appointment began
* ‘End’ – The UTC timestamp when the appointment ended
* ‘PID’ – The unique identifier representing the service provider
* ‘CID’ – The unique identifier representing the customer
* ‘TD’ – Length if the appointment
* ‘hour’ – hour of the timestamp when the meeting begins
* ‘month’ – month of the timestamp when the meeting begins
* ‘year’ – year of the timestamp when the meeting begins

Description of the Dataset

RangeIndex: 14080599 entries, 0 to 14080598

Data columns (total 9 columns):

start datetime64[ns]

End datetime64[ns]

PID float64

CID float64

TD timedelta64[ns]

hour int64

month int64

day int64

year int64

**Methods**

The data set is preprocessed for null values and then the time stamp attributes are converted to year, month and day as required for the analysis. All the attributes are grouped according to the ‘PID’ for analysis. After the preprocessing, the dataset is clean and ready for analysis.

**Initial Analysis of Data set**

No of appointments (based on appointment ID) = 14080599

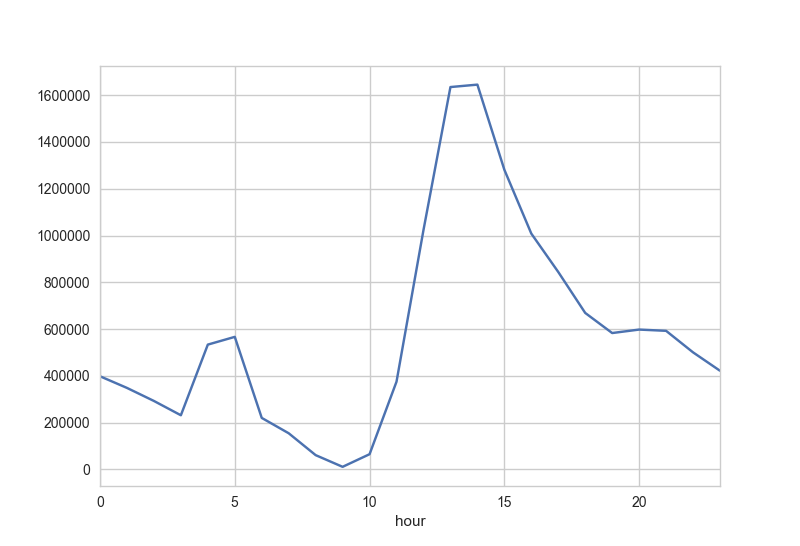
No of Customers (based on customer ID) = 179879

No of service providers (based on provider ID) = 137535

**Q1. How should ACME Corp think about its service provider staffing needs based on an**

**understanding of its current customer base**

The following are the outcomes of the Service Providers based on Customer data and timestamps.

1. *Time of day when most number of customers handled*.

**Fig 1: No of customers per hour**

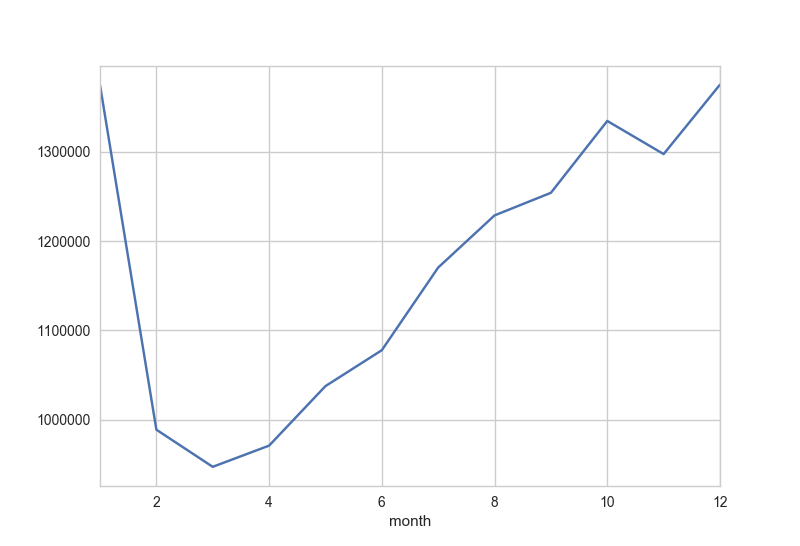
![C:\Users\Balaganesh\AppData\Local\Microsoft\Windows\INetCache\Content.Word\heat_map[day,hour](PID).png](data:image/png;base64,)

**Fig 2: The heatmap of the hour of day when the appointment begins**

**Ref: {'Mon',1:'Tue',2:'Wed',3:'Thu',4:'Fri',5:'Sat',6:'Sun'}**

* Most number of customers are handled by the providers in the afternoon.
* The ideal time when most of the appointments are set is between 1:00 pm – 3.30pm.
* The ACME corp. should consider appointing more number of service providers during the DAY shift. ( ideally 10am – 6 pm )

1. *Total number of requests handled in every month of all years*



**Fig 3: No of customers per month**

![C:\Users\Balaganesh\AppData\Local\Microsoft\Windows\INetCache\Content.Word\heat_map[month,day](PID).png](data:image/png;base64,)

**Fig 4: Heat map of the day of the month when the appointment begins**

**Ref: {'Mon',1:'Tue',2:'Wed',3:'Thu',4:'Fri',5:'Sat',6:'Sun'}**

* Most number of customers are handled during the 2nd ,3rd ,4th quarters of the year.
* The ideal time of the year when most of the appointments are set is in-between JULY -DECEMBER of each year.
* The most number of appointments set is the in months of OCT, NOV, DEC (4th Quarter)

Recommendation from the analysis

* The ACME corp. should appoint more number of service providers in-between JULY-NOVEMBER and should hire more number of service providers in-between OCT - DEC to serve more number of Customers.

**Q2 What appointment length is ideal considering a constrained pool of service providers and a normal eight‐hour workday**

1. *After observation of timestamps of all the years, below is the analysis of the data given*.

*Description of the TD (TIME DIFFERENCE) attribute*.

count 14080599

**mean 0 days 04:37:54.981054**

stnd 0 days 04:43:44.085738

min 0 days 00:00:00

25% 0 days 02:55:00

50% 0 days 04:00:00

75% 0 days 06:00:00

max 430 days 1

Average time per appointment for 1st quartile is ‘0 days 01:34:52.669383’

Average time per appointment for 2nd quartile is ‘0 days 03:25:08.716847’

Average time per appointment for 3rd quartile is ‘0 days 05:17:46.152983’

Average time per appointment for 4th quartile is ‘0 days 09:13:09.705868’

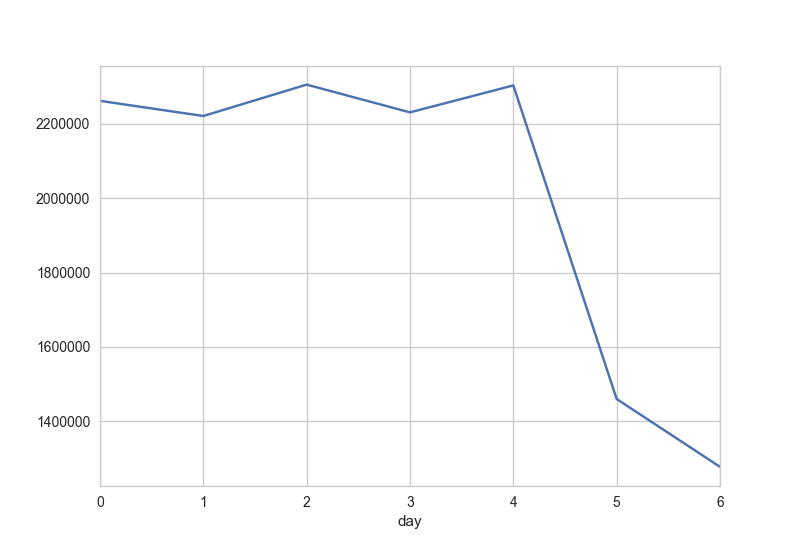
After the ideal observation of the data below are the metrics of the timestamps given.

* **Average time per appointment for all the quarters – 4.30 hours.**
* **After careful observation of each quarter and the analysis from the dataset, the ideal appointment length is 4.30 hours.**

**Additional Findings/ Recommendations**

After reviewing each aspect of time and count of the service providers, here are the additional findings and recommendation to the ACME corp.

1. *Days of the week when more number of appointments begin.*



**Fig 5: No of Customers per day in a week.**

**Ref: {'Mon',1:'Tue',2:'Wed',3:'Thu',4:'Fri',5:'Sat',6:'Sun'}**

* From the analysis using the data from all the years, we can conclude that more number of customers are starting their appointments during the 5 Day period. There are few appointments on the weekends too.
* The ACME corp. should appoint more number of service providers during WED to FRI

1. *Number of appointments per year.*

![C:\Users\Balaganesh\AppData\Local\Microsoft\Windows\INetCache\Content.Word\heat_map[year,month](PID).png](data:image/png;base64,)

Fig 6: Heat map of the year and month from the data

* From the data and analysis of each year, the number of appointment are increasing each year for ACME corp.
* Most number of appointments for ACME corp. was in the year 2016.
* ACME corp. can appoint more number of service providers to handle more number of customers.

1. The ideal time of the appointment is 4.30 hrs. in a normal 8-hour workday. The ACME corp. can improve its number of customers if the lenght of the appointment is cut to **3hrs – 2hrs /day.**

year: 2010, No of Appointment: 2

year: 2011, No of Appointment: 0

year: 2012, No of Appointment: 19

year: 2013, No of Appointment: 574

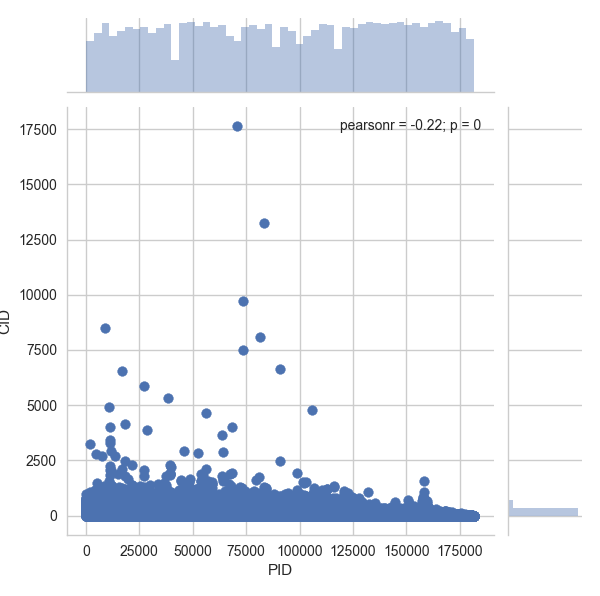
year: 2014, No of Appointment: 2090732

year: 2015, No of Appointment: 5270953

year: 2016, No of Appointment: 6037031

year: 2017, No of Appointment: 664775

**Additional Graphs**



**Fig 7: Joint plot of provider id vs customer id counts.**

* The above plot describes the number of customers handled by each service provider and the linear regression line is perfectly distributed equally for all the service provider.