MEDI-FIND

ETL ANALYSIS REPORT

INTRODUCTION

This report contains our analysis, regarding several hypothesis in relation to the data we have. This analysis is done by using the external data sources as follows :

EXTERNAL DATA SOURCES

1. Statistics of total number of memberships at fitness centers / health clubs in the U.S. from 2000 to 2015.

   https://www.statista.com/statistics/236123/us-fitness-center--health-club-memberships/

DATA

U.S. fitness center / health club memberships 2000-2015	Total number of memberships at fitness centers / health clubs in the U.S. from 2000 to 2015 (in millions)
2000	32.8
2001	33.8
2002	36.3
2003	39.4
2004	41.3
2005	41.3
2006	42.7
2007	46.7
2008	45.6
2009	45.3

2010	50.2
2011	51.4
2012	50.2
2013	52.9
2014	54.1
2015	55

2. Ranking of Insurance providers in US.

   http://health.usnews.com/health-news/health-insurance/articles/2013/12/16/top-health-insurance-companies

DATA

Insurance providers rank	Insurance provider
1	UnitedHealth Group
2	Kaiser Permanente
3	Wellpoint
4	Aetna Group
5	Humana Group
6	HCSC Group
7	Cigna
8	Highmark
9	Blue Cross and Blue Shield Association
10	Independence Blue Cross Group
11	Centene Corporation
12	HIP Insurance Group
13	Guidewell Mutual Holdings Group

14	Wellcare Health Plans
15	CareSource
16	Molina Healthcare
17	Lifetime Healthcare Group
18	Metropolitan Group
19	Cambia Health Solutions

ANALYSIS :

HYPOTHESIS-1 : (Related to Fitness)

If the number of memberships at fitness centers / health clubs are increasing over the years 2000- 2015 then the number of appointments will be decreasing over the same range of years.

ETL Workflow :

ETL Description :

This ETL workflow will read data from the existing tables in our database to get the number of appointments over the years 2000- 2015 (Use Aggregate to get the count) and the data from the external source for count of people who have started going to fitness centers.

Analysis Chart :

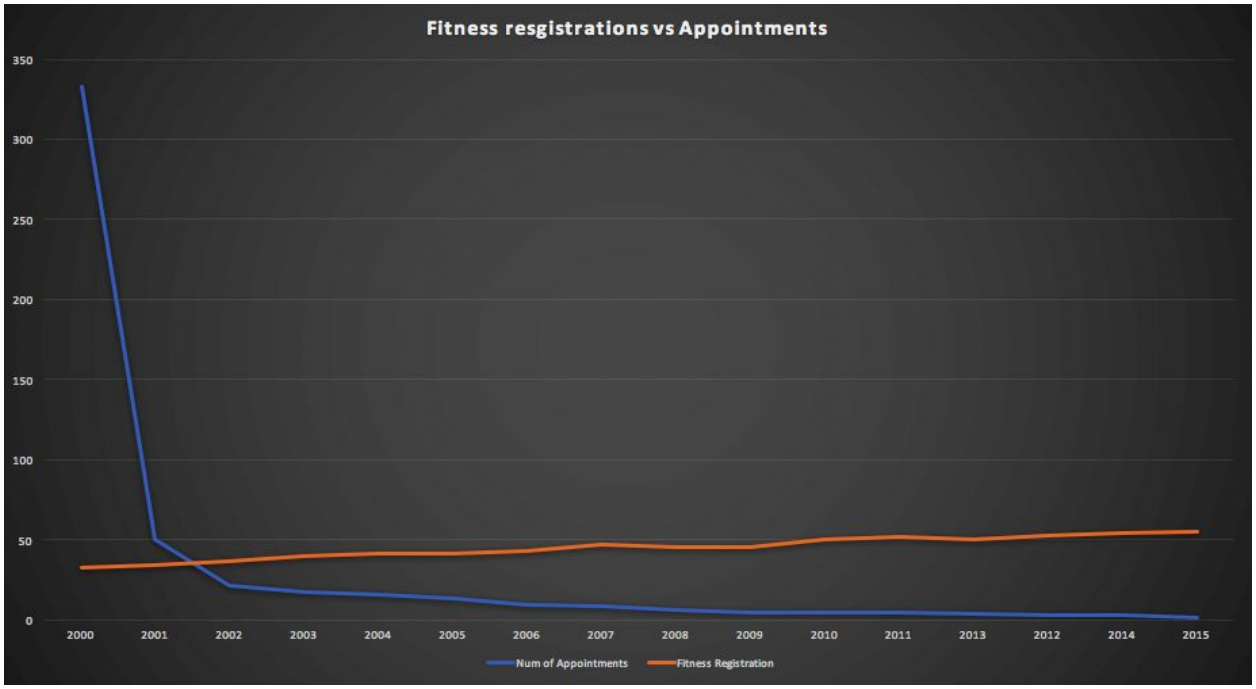


Chart Description :

X- axis : years over 2000 - 2015

Y- axis : shows the decline of number of appointments over years 2000- 2015 and increase in the number of people registering in fitness centers

Inference:

No of appointments have gone down with more number of people using fitness centers, hence validating the hypothesis.

Significance of the chart.

We know that the usage of our application to make appointments might reduce at times of the year when people typically sign up for fitness membership. We could maybe look at adding the functionality for app users to also join gyms so we can maintain users by offering appointment services and gym membership services.

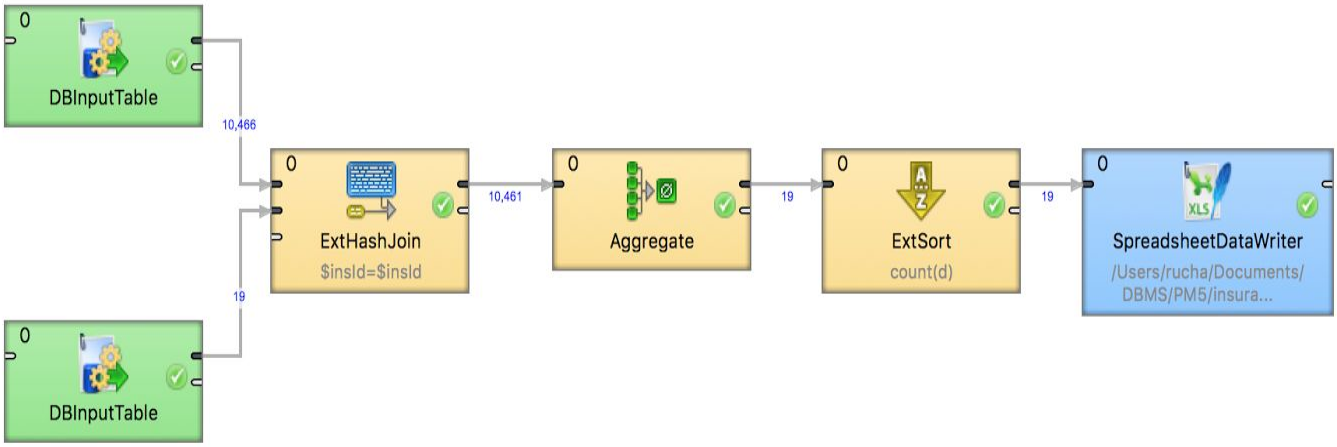
HYPOTHESIS-2 : (Related to ranking of Insurance providers)

1. More numbers of users are registered with insurance providers with higher

ranking.

2. More number of Hospitals associated with most ranked Insurance partners.
3. From the above statements we can state that there will be more number of appointments for healthcare facilities which supports highly ranked Insurance providers.

ETL Workflow_1 :



ETL Description :

1. Join two tables users and Insurance ranking (table created from external data source)
2. Aggregate on Insurance Id from ranking to find the count of users who uses the insurance providers listed in the ranking data source.
3. Sort the result based on the ranking.

Analysis Chart :

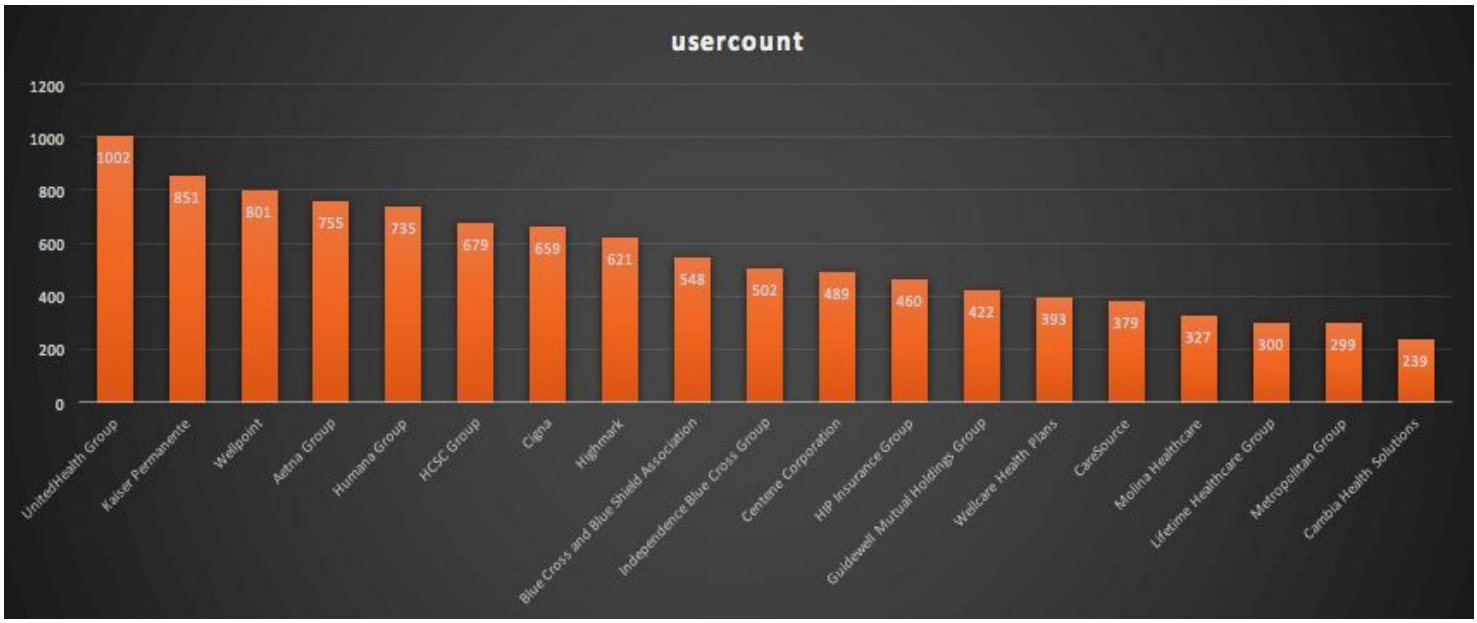


Chart Description :

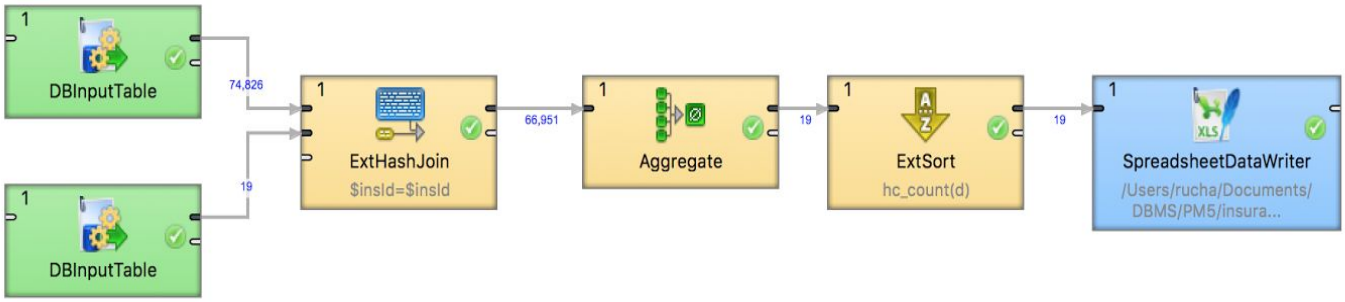
X- axis : Insurance providers according to the rank.

Y- axis : User count

Inference:

More number of people prefer to obtain healthcare insurance services with insurance providers that have the best rankings according to the analysis.

ETL Workflow 2 :



ETL Description :

4. Join two tables HealthcareFacility and Insurance ranking (table created from external data source)
5. Aggregate on Insurance Id from ranking to find the count of healthcare facilities who supports the insurance partners listed in the ranking data source.
6. Sort the result based on the ranking.

Analysis Chart :

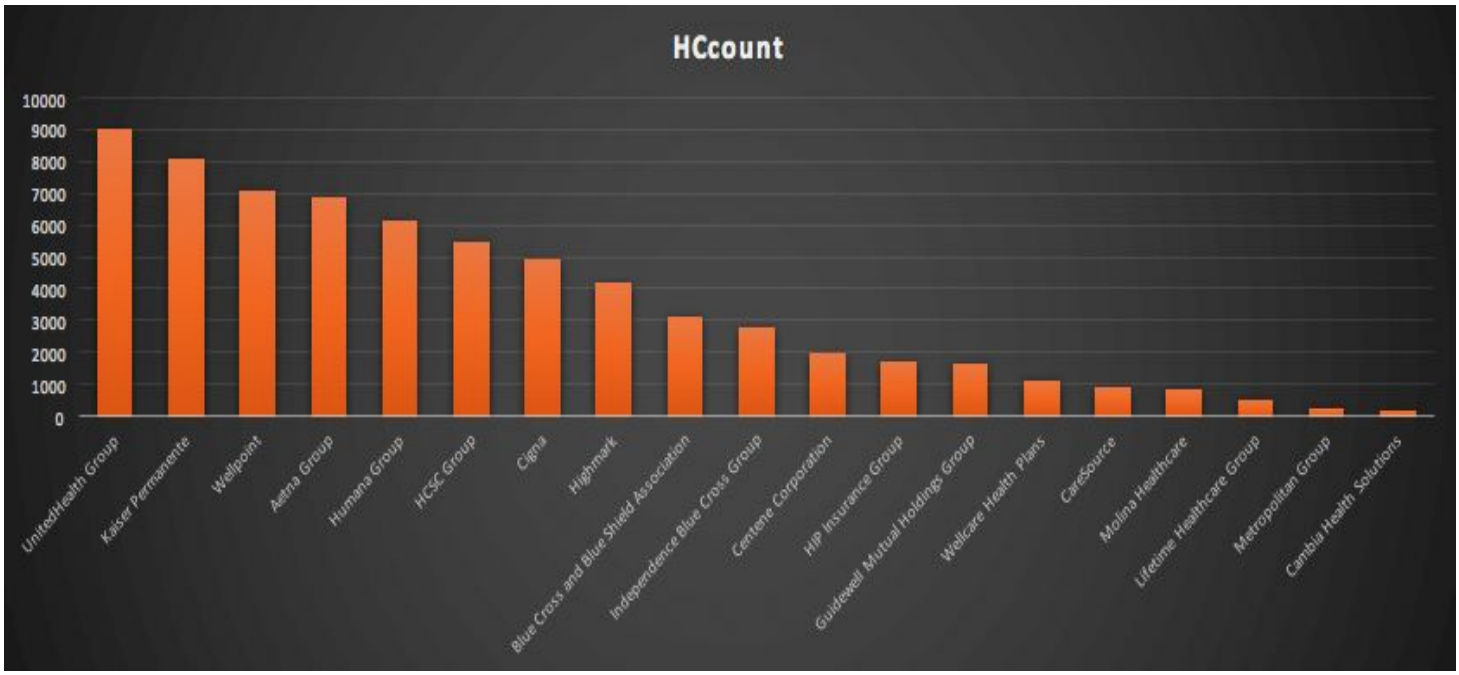


Chart Description :

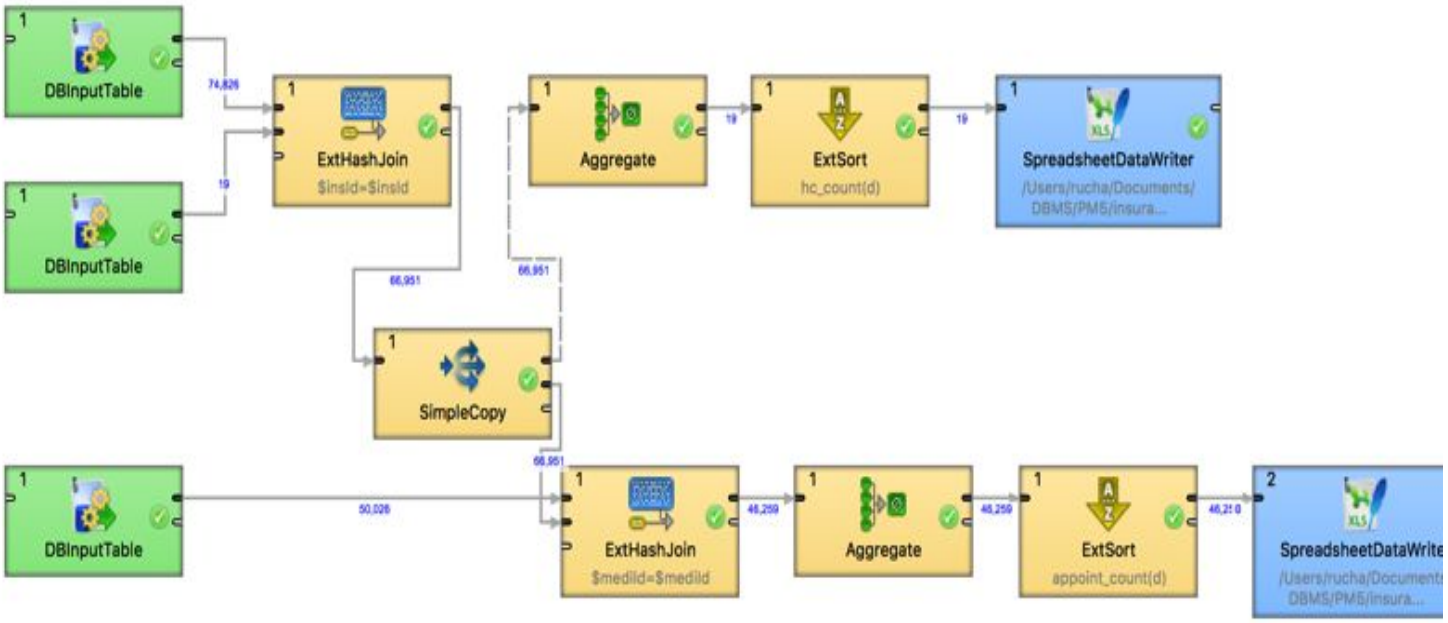
X- axis : Insurance providers according to the rank.

Y- axis : Healthcare Facility count

Inference:

More Healthcare facilities have partnered with Insurance providers which are highly ranked, which validates our hypothesis.

ETL Workflow 3 :



ETL Description :

This ETL makes simple copy of the previous workflow for calculating the number of appointments per insurance provider which is listed in the top ranked ones.

1) It makes an inner join with the previous results on the MediFacilityId
2) Then aggregate with the Insurance Name
3) Sort with the count of appointments
4) Write the results to an excel sheet.

Analysis Chart :

Chart Description :

X- axis : Insurance providers according to the rank.

Y- axis : Count of appointments (aggregating all healthcare facility which is associated with a specific insurance provider)

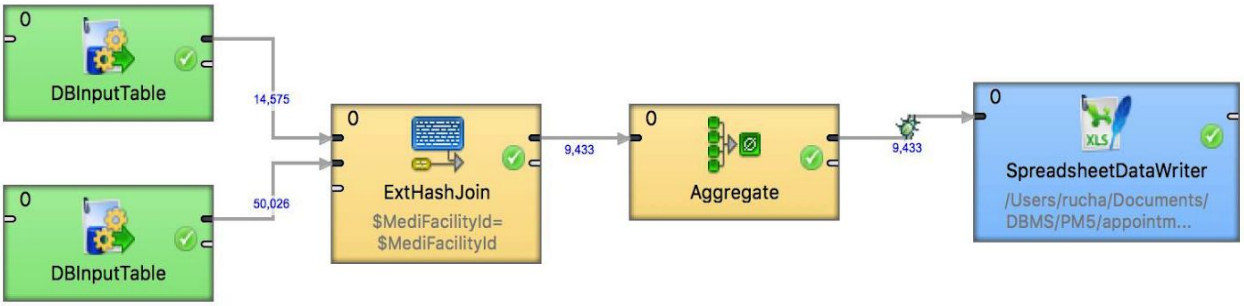
Inference:

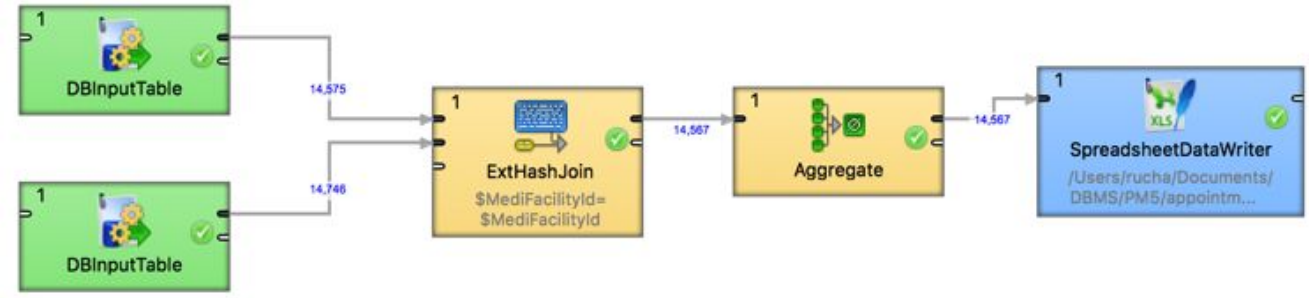
From the trend of the graph we see a higher number of appointments at the highly ranked healthcare facilities and this supports our initial hypothesis

HYPOTHESIS-3 : (Related to rating of Healthcare facilities)

Healthcare facilities which are highly rated in the system will be preferred by most number of users.

ETL Workflow :





ETL Description :

This ETL workflow will read data from the existing tables to get the number of appointments over the years 2000- 2015 (Use Aggregate to get the count) and the data from the external source for count of people who have started going to fitness centers.

Analysis Chart :

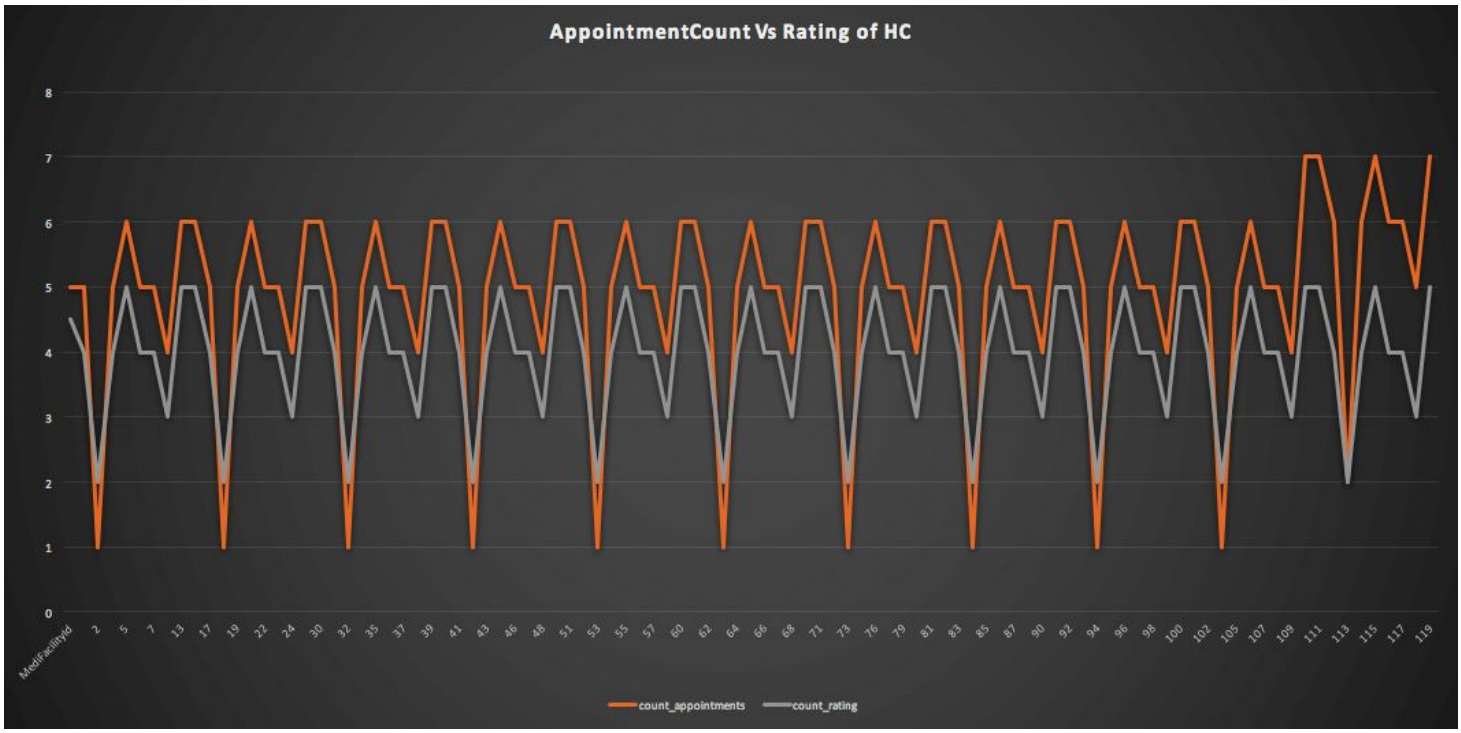


Chart Description :

X- axis : Identifiers for healthcare facilities

Y- axis : On Y-axis we represents the trend of ratings for specific healthcare facility(grey line) and the number of appointments for a specific facility.

Inference:

The trend is proportional for the number of appointments and the rating of the healthcare facility, hence validating the hypothesis that the a highly ranked healthcare facility has more appointments

Significance of this chart to our application.

We know that users are probably going to have the top insurance providers as there users so we could work with that provider to provider exposure on our app in turn for revenue sharing. Since we know there is a high probability users use these insurance providers, there must be a high probability that we would make revenue from those users.

CONCLUSION

Each of the hypothesis we put forward were validated using data we extracted and transformed using CloverETL.