WHAT ARE THE KEY DRIVERS OF PATIENT RETENTION?

Team 127: Lea Peters, Damilola Fadele, Friday James, Nick LaRosa, Chad Morton, Mohamed Toure

Highlights

- Ensuring that patients are not only receiving the best possible treatment but are also happy with staff interactions, and overall experience are equally important to patient retention.
- Understanding patient factors and demographics as they relate to retention among new patients allows for clearer navigation through post-pandemic hurdles.

Background

Embracing and improving patient retention has always been an important aspect of cancer research, trial, and treatment.

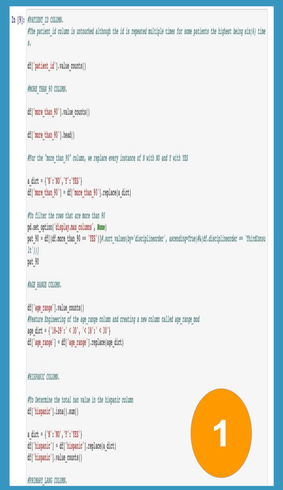
Our sponsor is interested in understanding patient factors and demographics as they relate to optimizing patient retention, specifically regarding new patients

We hope to aid in improving their patient retention strategies by identifying motivating factors or patient satisfaction.

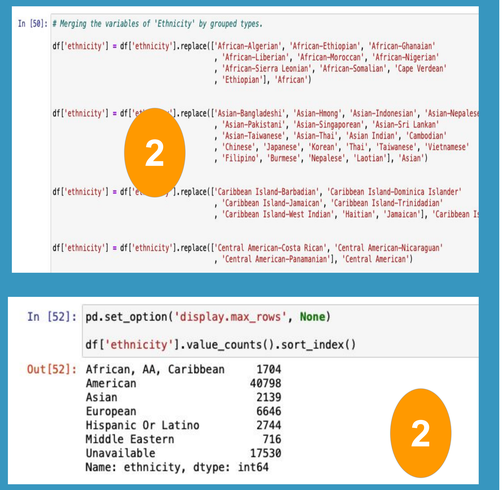
Data Cleaning / Engineering

Dataset Cleansing Method: Our team standardized data types and column headers within the dataset to make it a manageable structure.

Feature Engineering: Our team created new groupings of the data in order to make them easier to interpret and model



1



2

```
In [52]: pd.set_option('display.max_rows', None)

         df['ethnicity'].value_counts().sort_index()

Out[52]: African, AA, Caribbean     1704
         American                  40798
         Asian                      2139
         European                   6646
         Hispanic Or Latino         2744
         Middle Eastern              716
         Unavailable               17530
         Name: ethnicity, dtype: int64
```

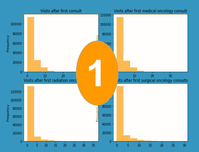
2

Exploratory Data Analysis

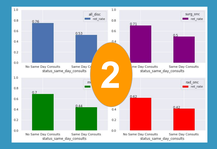
Univariate analysis on key variables in our dataset such as the number of visits after first consult.

Bivariate analysis to explore the relationship between our four retention variables (overall, medical oncology, radiation oncology, and surgical oncology) and several factors such as: age, gender, race, type of insurance, type of disease (cancer), discipline count, modality of first consult, and primary language

Multivariate analysis to explore additional interactions between our variables



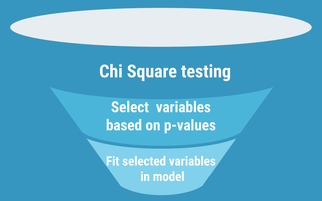
1



2

3

Modeling

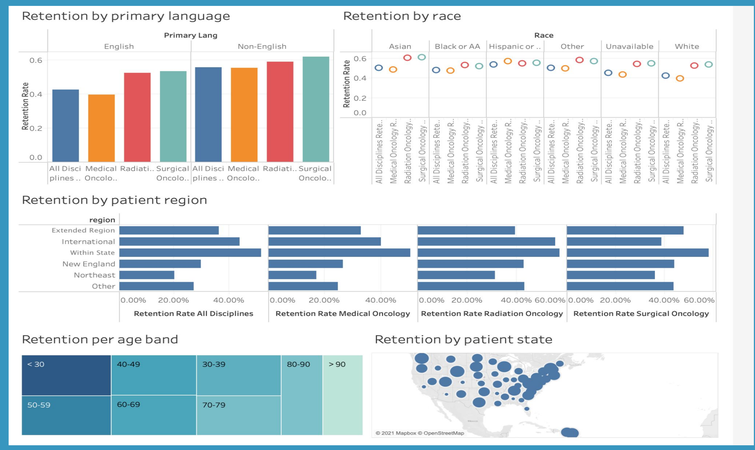


LOGISTIC REGRESSION

Chi-Square analysis was performed to determine the association between the features and the four retention variables. Majority of the features were correlated with retention, thus a subset of them were selected based on the highest p-values from the analysis. The selected features were fitted into a logistic regression model

Dashboard

Our dashboard displays the various conclusions and findings we observed when analyzing the dataset. There are various charts, maps, and figures that visually describe our findings and are grouped by category for easy navigation.



Conclusions

The results from the logistic regression model were in conformity with that of our exploratory data analysis

Future Work: As telemedicine continues to gain popularity, it would be interesting to see how this impacts retention especially for patients with residences out of state

Additionally, as more data is collected on two recently piloted programs: patient coordinator and nurse calls prior to consults and transportation services for patients, it would be interesting to see how it impacts retention