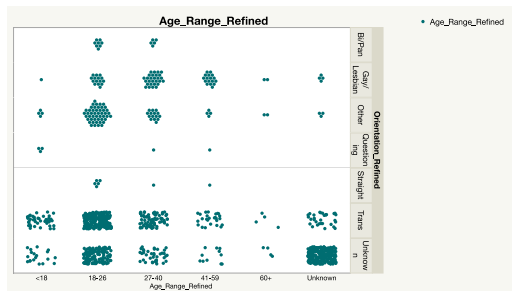**Objective:**
- Provide better care for the patients.
- Improve efficiency of the clinic operations

**Drawback:** Patient-generated health data. Therefore, the data is quite unstructured.
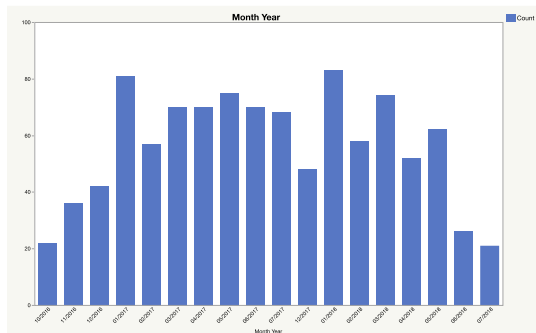
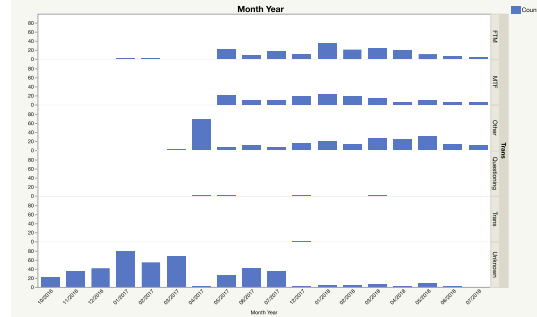**Data Cleaning and Data Visualization**
**Tools:** Python, SAS JMP
**Github:** https://github.com/aviroop1/LGBT-Clinic-Data-Analysis



*Fig 1: Age Range vs Orientation*

- Majority of the visitors belong to the age group 18-26.
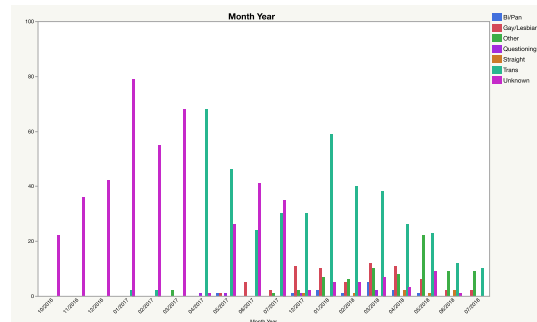- Within age group 18-26, there are significant number of transgenders and Gay/Lesbians



*Fig 2: Count of visitors per Month Year*

- January is the busiest month
- Friday is the busiest day of the week
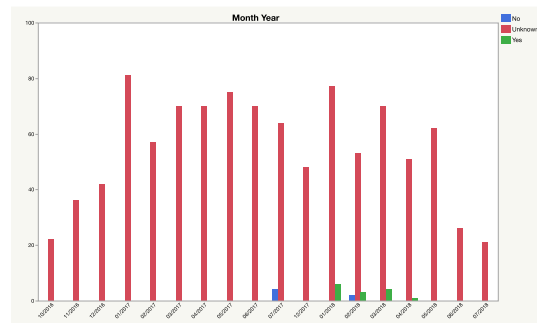- Jan 2017 to May 2018 – There was a high arrival of visitors in the clinic



*Fig 3: Count of Visitors per Month year grouped by Transgender category*



*Fig 4: Count of Visitors per Month year grouped by orientation*

FTM, MTF and Gay/Lesbians started visiting the clinic since May 2017. Initially the information of the visitor was unknown. Either the data wasn't collected effectively or due to social stigma, patients were embarrassed to give more information.
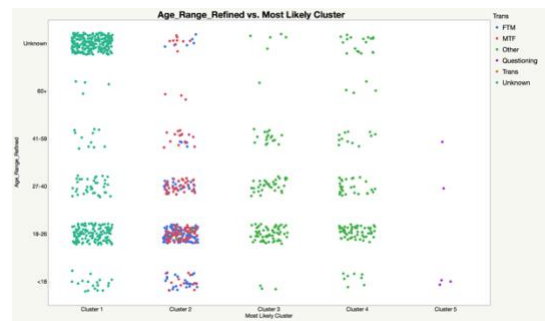


*Fig 5: Count of visitors per month year by HIV positive patients*

- Jan to Mar 2018 witnessed HIV positive visitors
- May 2017 witnessed HIV negative visitors

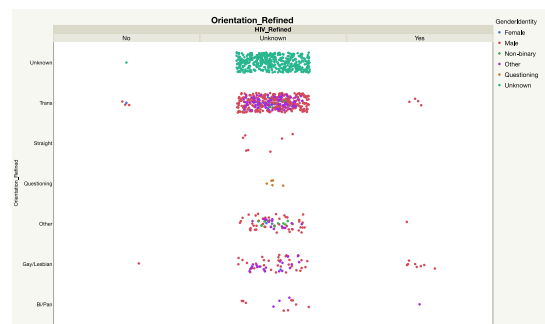**Unsupervised learning: Latent Class Analysis**
**Tool used: SAS JMP**

Using Latent Class Analysis, the population was categorized into 5 different clusters using features Age Range, Trans, Orientation, Gender Identity, HIV, Prep.
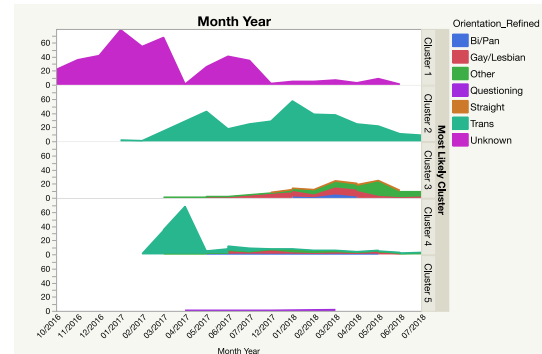


- The Cluster 2 and Cluster 4 mainly comprise of Transgenders.
- The Cluster 2 comprises of MTF' and 'FTM'. Majority of them are part of age groups 18-26 and 27-40



Cluster 2 and Cluster 3 comprise of HIV infected patients.



HIV positive victims mainly comprise of males, who are Gay/Lesbian or transgenders.



The above graph represents the arrival pattern of the different clusters among the patients that visited the clinic. The 'MTF' and 'FTM' patients have always been visiting the clinic. In the last 2 months, the number of patients visiting their clinic has gone down.

**RECOMMENDATIONS:**

1. **Recommendation A**: Clinic can also identify 4 categories of patients visiting the clinic. These 5 categories comprise of:
   a) Patients who have not given us sufficient information
   b) Transgenders (MTF and FTM), whose age range is from 18-40, HIV infected victims
   c) HIV infected victims: Gay/Lesbians, Bisexual
   d) Transgenders
2. **Recommendation A:** Since transgenders and Gay/Lesbians are starting to get rid of social stigma (starting May 2017), the clinic can extract more data from the LGBT patients, hence ask more questions on the LGBT status and interaction with others due to which they are HIV infected or are taking PREP, and list out causes of the issue.
3. **Recommendation B:** The survey or the questionnaire should have *multiple choice questions* and not empty answer boxes, where patients have the freedom to write anything.
4. **Recommendation C:** In the beginning of the year, the clinic should ensure that they optimize their service due to arrival of high number of patients.