**Meeting Notes for November 3:**

Agenda: Updates on findings/progress from last week

Mei:

Choosing suitable minsup

2 methods:

* Apply FP-growth across different category (by Type or by Neighbourhood)
  + Use median to calculate minsup (median requests/ total requests)
* Or use adjustable minsup formula according to datasize
* Will have to consult with Carson for advice

**Tasks:**

* **Consult with Carson regarding the most appropriate method to choose minsup over Discord. (we can advise on the wording of the message before you send it)**
* **Begin draft on the final report**
  + **Include sections**
  + **Can fill out introduction, contribution, and the methodology of selection of minsup**

Shika:

* The final data will be frequent itemsets, combination of neighbourhood, ward, type and reason
* If we’re breaking it down by neighbourhood, the itemsets will include ward, type and reason
* If we’re breaking it down by type, the itemsets will include neighbourhood, ward and reason
* Reason for doing this(separating/breaking it down by different neighbourhood or type) : Uneven distribution of request frequency across these categories, which will affect the minsup calculation we use
* Anomaly data
* So things like Pie charts, line charts, mapping etc.(do frequent for now, leave anomaly for later)
* Use columns Reason, Type, Neighbourhood, Ward as the columns. The other columns are not relevant
* Two columns (itemset and support) - for final

**Tasks:**

* **Decide on which initial data would be good to visualize for both report and presentation, include reasons why they would be good, and make those visualizations**
* **Begin preparing to visualize frequent itemsets that we find (see above for more details on what the itemsets will be)**
  + **This would be in the form of code or steps in a software such as Tableau**

Aiden:

* To detect anomalies, will have to define normal patterns
* Will apply time-series + clustering to determine anomalies
* Datapoints far from cluster are anomalous
* Time periods will be seasonal or yearly
* What we’re looking for:
  + General: abnormal frequency of request type/reason in given time period, based off of past data
  + Within a neighbourhood: abnormal frequency of request type/reason in given time period, based off of past data
* Look into how to transform data, compare with examples

**Tasks:**

* **Make a plan for implementing anomaly detection (what algorithm & library to use, what data transformation/cleaning needed), and begin implementation**

Kelvin:

* Cleaned and formatted the data for testing the FP-growth algorithm from the SPMF library, using Python scripts (<https://github.com/yenaing-oo/COMP-4710-Project>)
* It works! Will have to decide on suitable minsup, and also whether/how we’re categorizing the data by neighbourhood and/or type
* Will format the results into a csv file for Shika to work with for visualization

**Tasks:**

* **Look into how to write analysis for frequent itemsets by studying other papers**
* **Begin writing analysis once frequent itemsets are gathered (blocked by selection of minsup and whether to apply categorization)**

**Next meeting: Sunday Nov 10 at 2pm**