# Meeting Minutes - Thur Nov 4th 2021

Attendees: Esther, Jerry, Wisam, Alec, Olesya

## Notes

* Look for demand in industries to look for what people are looking for and tie it to the people working.
* Looking for historical data from stats canada
  + Job vacancies over the past few years
* Goal - Which industry is at risk based on the amount of people retiring.
* Two datasets required
  + Multiple years for Census and multiple years of job listing, the years don’t have to match we can use machine learning to fill in missing trend data if job listing does not contain matching years data.

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610048901

* Shawn said to find census Canada job listing data.
* Machine learning inputs, Age, Salary, Industry
  + Model: Find Industry retirement age, if Age > Industry retirement age + or - 3 then at risk of leaving.
  + Output: at Risk or not at Risk
  + We then take sum of counts of risk value per industry to show industry is at risk for losing multiple people.
* Salary could be a sign if someone would retire soon or not?
  + Thoughts on why: High Salary CEO would generally have enough money to retire early but it does not mean they will, low playing employees might need more money so might keep working past average retirement age.
* Industry, physical vs mental, Physical industry most likely will have a lower retirement age as it is too tasking for the body, while mental some people will stay in the industry and be useless and collect money.

**TASKS NEEDED TO BE DONE BY SUNDAY**

1. Create Mock up of machine learning model
2. Create ERD mock up
3. Create an outline for final dashboard
4. Create branches on GitHub with working code for EDA

EDA - Exploratory data analysis

**DATA SET found by Jerry**

**https://open.canada.ca/data/en/dataset/67f90ff0-12ea-429a-99a6-7b41c73863a0/resource/2cfa6a73-0b66-4b6e-a07b-7285f0ea774c**

## Questions for Shawn

* Can upload large sets of data to tableau
* Issue is figuring out topic
* Arima model or neuro network model to forecast the employment rate of certain industries.
  + Could be a regression problem
  + Use the employment rate of a previous time period to determine the employment rate of the next time period

Time series can check if something is working or not.

Good linear regression model

HASSAN

Diagram and use text to explain models, what is output of classification of model, what are the inputs of the model, within powerpoint, be ready to explain each component

Machine learning should not be python working code, be sure in your own environments be ready to start, only flow diagram,

ERD, CREATE STATEMENTS AND SAMPLES FOR EACH TABLE AND MOCK UP OF MACHINE LEARNING

**OUR TOPIC**

**Filtering on Job vacancy characteristics it has redundancies, trying to find highest or lowest level of granularity**

**Topic**

**Potential number of vacancies in the next quarter by NOC.**

**Questions**

* **Prediction on Hourly wage**
* **Machine learning model** 
  + **Input:**
  + **Output:**

## Segment 1 - Deliverables

### Presentation

| **Task** | **Assigned to** | **Due** |
| --- | --- | --- |
| Editing ***ReadMe.md*** to state:   * Selected topic * Why this topic was selected * Describing the data source * Questions they hope to answer with the data (the goal of the project) |  |  |

### Github

| **Task** | **Assigned to** | **Due** |
| --- | --- | --- |
| Editing ***ReadMe.md*** to state:   * Description of the communication protocols |  | Nov 7, 2021 |
|  |  |  |
| Create individual branch   * Ensure 4 commits | Esther | Nov 7, 2021 |
| Create individual branch   * Ensure 4 commits | Wisam | Nov 7, 2021 |
| Create individual branch   * Ensure 4 commits | Jerry | Nov 7, 2021 |
| Create individual branch   * Ensure 4 commits | Olesya | Nov 7, 2021 |
| Create individual branch   * Ensure 4 commits | Alec | Nov 7, 2021 |

### Machine Learning

| **Task** | **Assigned to** | **Due** |
| --- | --- | --- |
| Provisional machine learning model  (a stand in for the actual final model)   * Takes in provisional database * Outputs label(s) for input data   *\*NOTE: this should probably be done by at least 2 people* |  | Nov 7, 2021 |

### Database

| **Task** | **Assigned to** | **Due** |
| --- | --- | --- |
| Create ERD of Datasets |  | Nov 7, 2021 |
| Provisional database  (a stand in for the final database)   * Sample data that mimics the expected final database structure or schema * Draft machine learning module is connected to the provisional database   *\*NOTE: this should probably be done by at least 2 people* |  |  |

### Dashboard

| **Task** | **Assigned to** | **Due** |
| --- | --- | --- |
| Outline for final Dashboard   * Where are we hosting the dashboard? * What is mock lay out and what kind of data do we * want to show the client? * What kind of story do we want to tell? |  | Nov 7, 2021 |

### 

.