DAB302- Health Care Analytics: Final Project Report: Section: 2 - Group:12

Health Surveillance System

Team Members: Alok Suresh Chilka

Pallavi Ravikumar Menon

Sonia Rajput

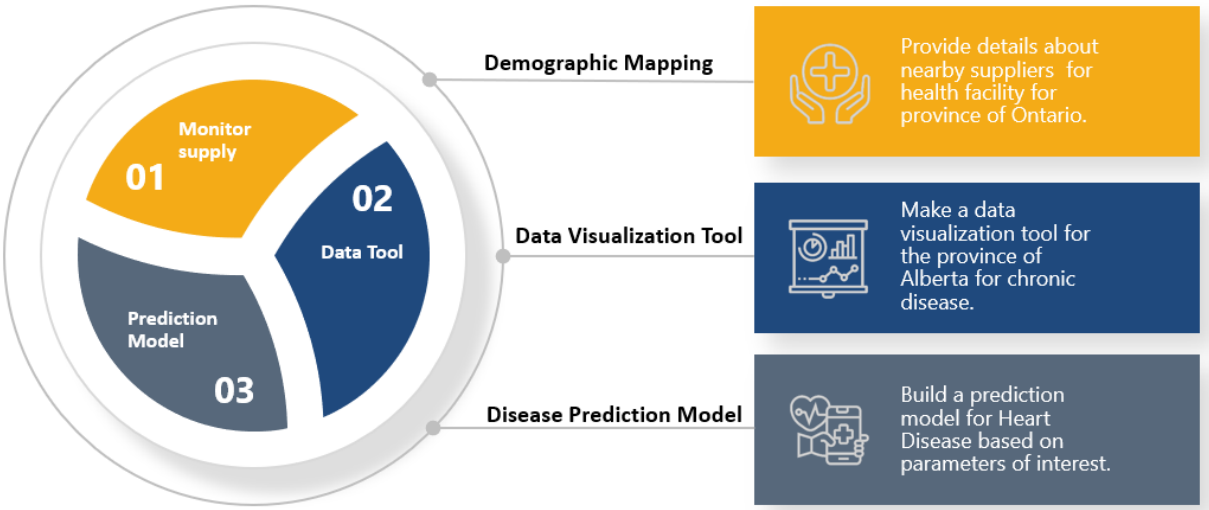
Abstract

Data is the fuel that drive decisions for tomorrow. The impact of analytics in healthcare domain can benefit in investigating early identification of diseases and enhance the infrastructure of traditional healthcare. The objective of our project is to provide a multidisciplinary environment for the hospital administration to identify, monitor and control prediction of chronic diseases majorly and for the administration to get an insight of the patterns of various diseases across the specific Canadian province. We also plan to incorporate channel that helps the hospital to the search various medical equipment to timely ensure availability of the products thereby ensuring best possible healthcare experience. Also, a demographic that visualizes various health care providers across the province. This project would be a tool to predict the health status and behavior of the populations concerned to chronic diseases based on parameters of interest. The purpose of surveillance is to empower decision makers to lead and manage more effectively by providing timely, useful evidence.

Contents

Introduction:

Health Surveillance is about systematic collection, analysis, and interpretation of data, closely integrated with the timely dissemination of these data to those responsible for preventing and controlling disease and injury. This project would be a tool to predict and visualize the health status and behavior of the populations. The purpose of surveillance is to empower decision makers to lead and manage more effectively by providing timely, useful evidence. There are several datasets for diseases with respect to different provinces starting from year 2000 until 2016. So, all datasets would be merged to create some meaningful visualizations out of it.



Dataset:

We have used datasets from different sources as our objectives contribute to different purposes:

- The supplier data is restricted to Ontario province for selected medical equipment.
- For the overview and analysis of chronic diseases we have data from the Alberta health wherein we can generate and customize visualization that could give deeper understanding for the patterns in monitoring certain diseases throughout the years/decade.
- Demographic dataset for various health provider locations for Ontario Province.
- The dataset used is from an open database for heart diseases. A comparative study was done with two separate Machine Learning algorithms using Logistic Regression and Tabnet Neural classifier. The software tool used was Jupyter Python notebook.

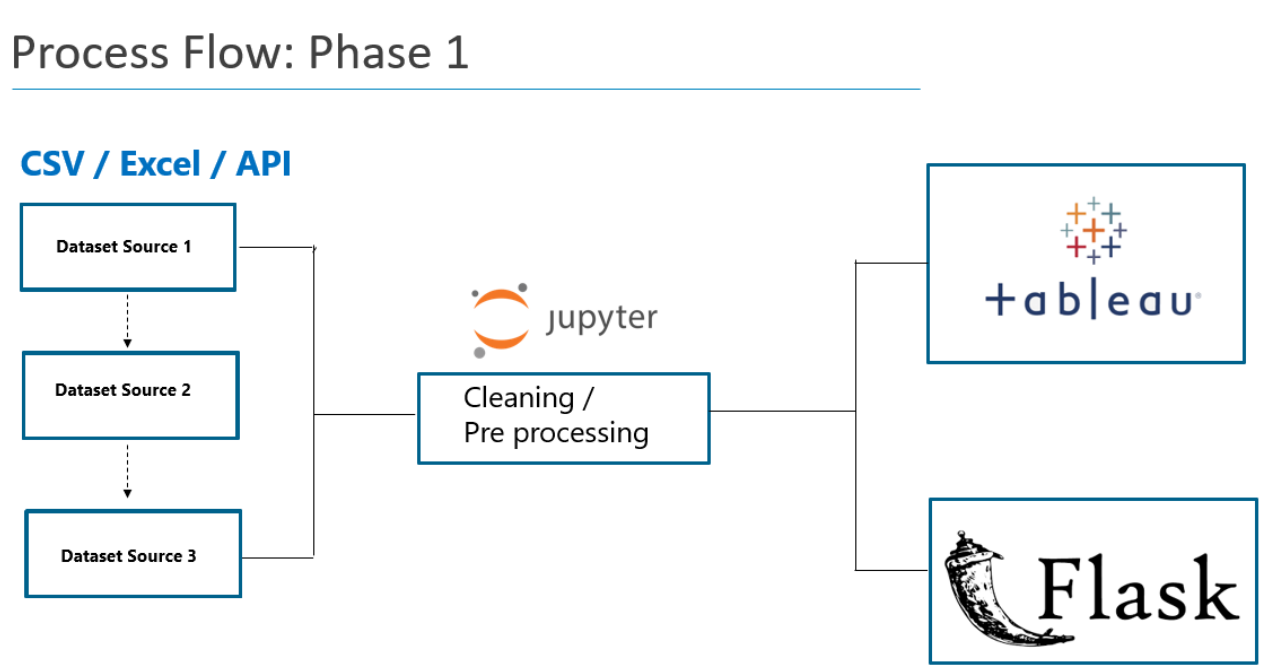
- Out of various chronic diseases, only some were selected based on Age-Sex Standard Prevalence.

- The tabular visuals represent the datasets used and the chronic datasets that were selected for the project.

Sr No.	Name of Chronic Disease	Age-sex Standardized Prevalance
1	Asthama	
2	Atrial Fibrillation	
3	Congestive Heart Failure	
4	Diabetes	
5	Hypertension	
6	Ischemic Heart Disease	
7	Liver Cirrhosis	

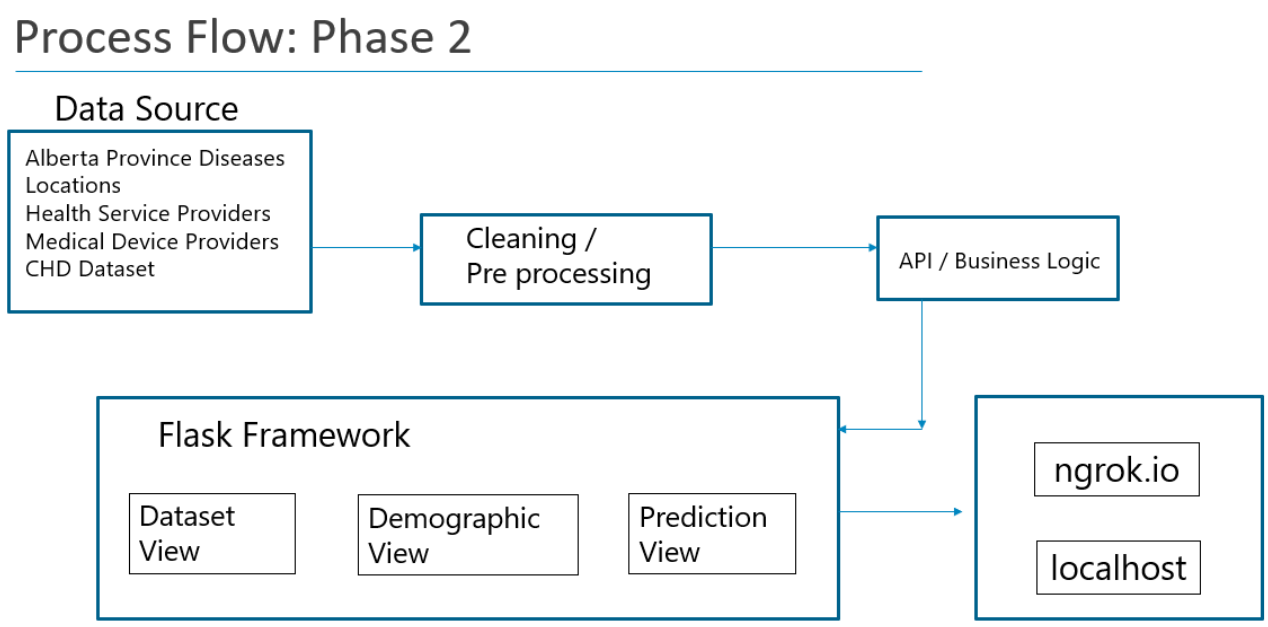
Sr.No		Purpose	Region	Category
1	Garments, pumps, and braces vendors	Supply	Ontario	Supplier
2	Home oxygen therapy vendors	Supply	Ontario	Supplier
3	Mobility aid vendors	Supply	Ontario	Supplier
4	Facial prosthetics vendors	Supply	Ontario	Supplier
5	Prosthetic limbs vendors	Supply	Ontario	Supplier
6	Respiratory equipment and supplies vendors	Supply	Ontario	Supplier
7	Communication aid vendors	Supply	Ontario	Supplier
8	Visual aid services	Supply	Ontario	Supplier
9	Ministry of Health service provider locations	Health service provider locations demographics	Ontario	Health Facilities
14	Alberta health Data	Visualization of various diseases	Alberta	Disease statistics
15	Chronic Disease Model	Prediction Model		ML Model

Project Workflow:



Process Flow: Phase 1

- The dataset included multiple csv's, hence the cleaning and preprocessing was done in Jupyter notebook for all the datasets.
- The initial phase included Exploratory data analysis and extensive cleaning which was later used for visualization.

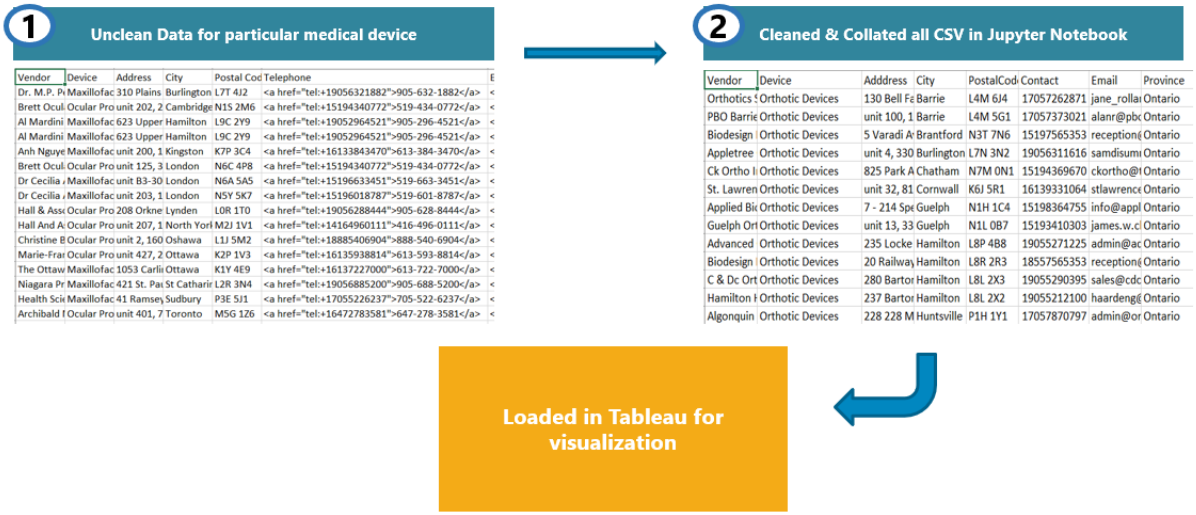


Process Flow: Phase 2

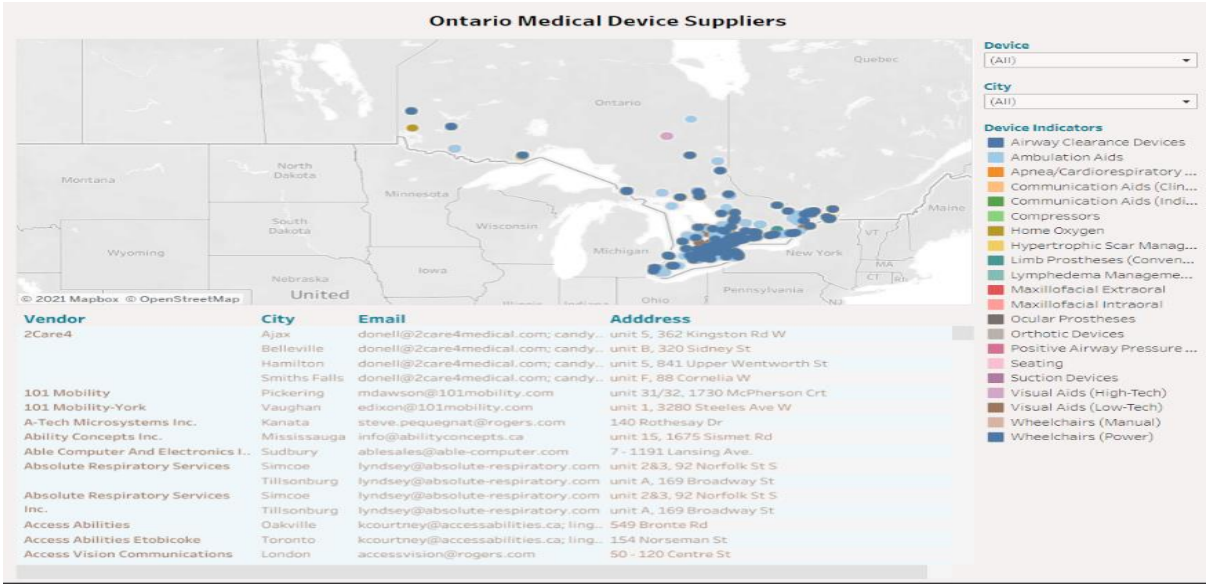
- The model is accessing the data via API. This is just read operation.
- The model results are simply displayed using parameters and graph plots.
- The application is hosted locally. Each dataset has different view.
- There are demographic views also to show the location of medical equipment and services.
- Model view shows in detail view of the steps for prediction.
- Flat file data format is being used in this model.
- Every dataset is passed through a cleaning phase to ensure there is no null / ambiguous data present.

Execution:

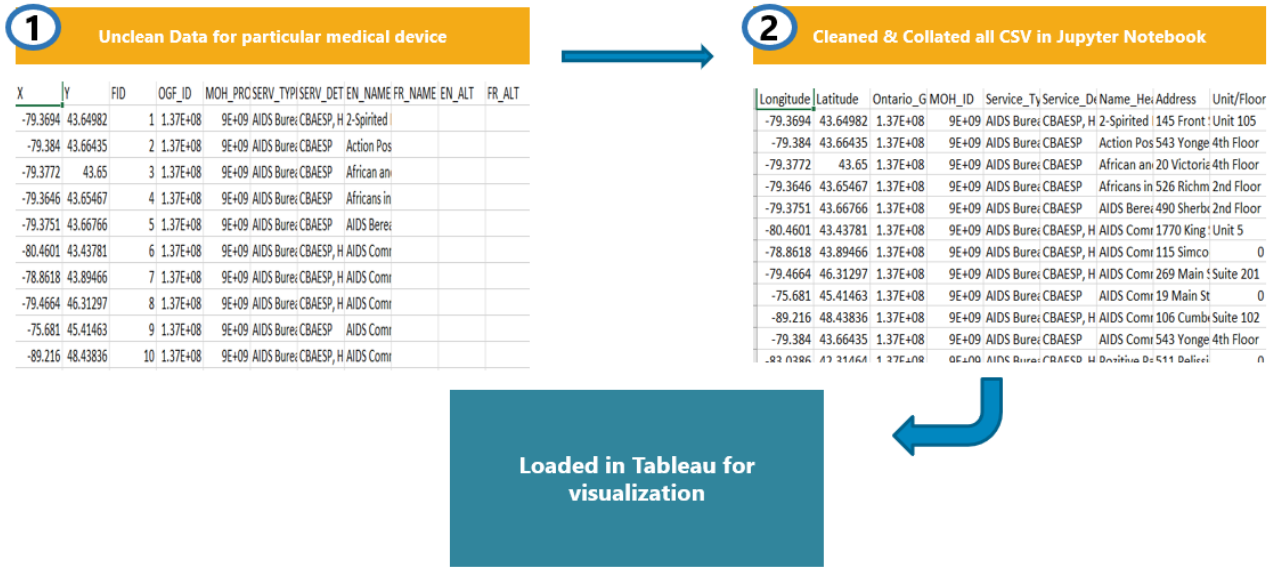
I. Supplier Dataset:



- The datasets were merged for various vendors of medical suppliers in Jupyter notebook.
- The figure above represents the uncleaned and cleaned supplier data, which was further incorporated into tableau.
- The figure below depicts the visualization of the supplier dataset, and we can observe the real time implementation in tableau using the various filters. The geographic representation for various suppliers is dynamic and is dependent on the applied filter.
- Furthermore, the visualizations were integrated into “Health Surveillance” API framework.



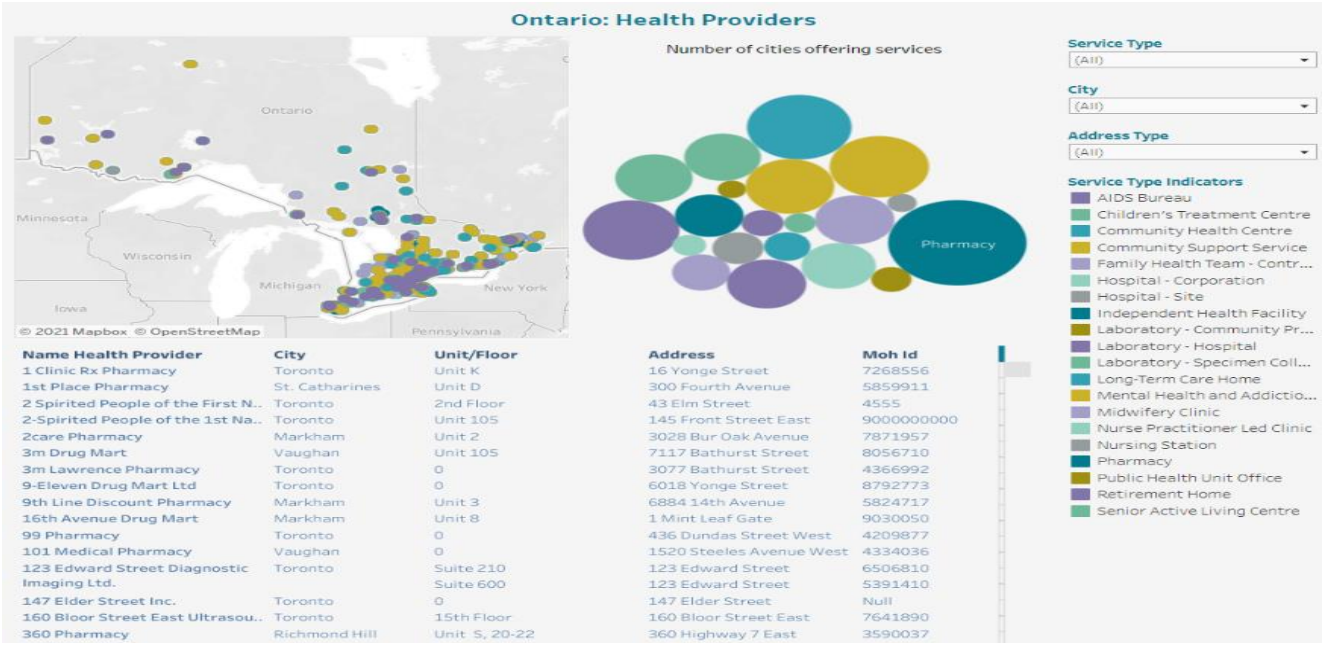
II. Health Providers Location Dataset:



X	Y	FID	OGF_ID	MOH_PRO	SERV_TYP	SERV_DET	EN_NAME	FR_NAME	EN_ALT	FR_ALT
-79.3694	43.64982	1	1.37E+08	9E+09	AIDS Bure	CBAESP, H	2-Spirited			
-79.384	43.66435	2	1.37E+08	9E+09	AIDS Bure	CBAESP	Action Pos			
-79.3772	43.65	3	1.37E+08	9E+09	AIDS Bure	CBAESP	African an			
-79.3646	43.65467	4	1.37E+08	9E+09	AIDS Bure	CBAESP	Africans in			
-79.3751	43.66766	5	1.37E+08	9E+09	AIDS Bure	CBAESP	AIDS Bere			
-80.4601	43.43781	6	1.37E+08	9E+09	AIDS Bure	CBAESP, H	AIDS Comr			
-78.8618	43.89466	7	1.37E+08	9E+09	AIDS Bure	CBAESP, H	AIDS Comr			
-79.4664	46.31297	8	1.37E+08	9E+09	AIDS Bure	CBAESP, H	AIDS Comr			
-75.681	45.41463	9	1.37E+08	9E+09	AIDS Bure	CBAESP	AIDS Comr			
-89.216	48.43836	10	1.37E+08	9E+09	AIDS Bure	CBAESP, H	AIDS Comr			

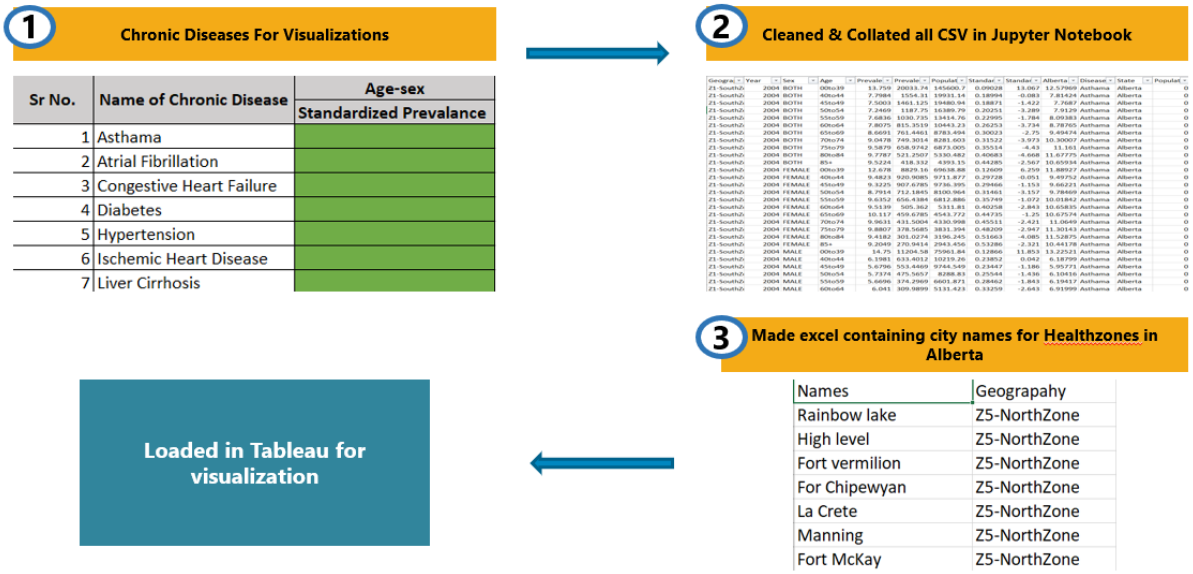
Longitude	Latitude	Ontario_G	MOH_ID	Service_Ty	Service_De	Name_Hea	Address	Unit/Floor
-79.3694	43.64982	1.37E+08	9E+09	AIDS Bure	CBAESP, H	2-Spirited	145 Front	Unit 105
-79.384	43.66435	1.37E+08	9E+09	AIDS Bure	CBAESP	Action Pos	543 Yonge	4th Floor
-79.3772	43.65	1.37E+08	9E+09	AIDS Bure	CBAESP	African an	20 Victoria	4th Floor
-79.3646	43.65467	1.37E+08	9E+09	AIDS Bure	CBAESP	Africans in	526 Richm	2nd Floor
-79.3751	43.66766	1.37E+08	9E+09	AIDS Bure	CBAESP	AIDS Bere	490 Sherbo	2nd Floor
-80.4601	43.43781	1.37E+08	9E+09	AIDS Bure	CBAESP, H	AIDS Comr	1770 King	Unit 5
-78.8618	43.89466	1.37E+08	9E+09	AIDS Bure	CBAESP, H	AIDS Comr	115 Simco	0
-79.4664	46.31297	1.37E+08	9E+09	AIDS Bure	CBAESP, H	AIDS Comr	269 Main S	Suite 201
-75.681	45.41463	1.37E+08	9E+09	AIDS Bure	CBAESP	AIDS Comr	19 Main St	0
-89.216	48.43836	1.37E+08	9E+09	AIDS Bure	CBAESP, H	AIDS Comr	106 Cumb	Suite 102
-79.384	43.66435	1.37E+08	9E+09	AIDS Bure	CBAESP	AIDS Comr	543 Yonge	4th Floor

- The figure above represents the uncleaned and cleaned health provider locations dataset, which was further incorporated into tableau for visualizations.
- The figure below depicts the visualization of the various health providers, by various service type.
- The dashboard lets the user filter down thee search type depending on the service type of various health providers as well as by the location or nearest city of the health units.
- Furthermore, the visualizations were integrated into “Health Surveillance” API framework.



Name Health Provider	City	Unit/Floor	Address	Moh Id
1 Clinic Rx Pharmacy	Toronto	Unit K	16 Yonge Street	7268556
1st Place Pharmacy	St. Catharines	Unit D	300 Fourth Avenue	5859911
2 Spirited People of the First N..	Toronto	2nd Floor	43 Elm Street	4555
2-Spirited People of the 1st Na..	Toronto	Unit 105	145 Front Street East	9000000000
2care Pharmacy	Markham	Unit 2	3028 Bur Oak Avenue	7871957
3m Drug Mart	Vaughan	Unit 105	7117 Bathurst Street	8056710
3m Lawrence Pharmacy	Toronto	0	3077 Bathurst Street	4366992
9-Eleven Drug Mart Ltd	Toronto	0	6018 Yonge Street	8792773
9th Line Discount Pharmacy	Markham	Unit 3	6884 14th Avenue	5824717
16th Avenue Drug Mart	Markham	Unit 8	1 Mint Leaf Gate	9030050
99 Pharmacy	Toronto	0	436 Dundas Street West	4209877
101 Medical Pharmacy	Vaughan	0	1520 Steeles Avenue West	4334036
123 Edward Street Diagnostic Imaging Ltd.	Toronto	Suite 210 Suite 600	123 Edward Street 123 Edward Street	6506810 5391410
147 Elder Street Inc.	Toronto	0	147 Elder Street	Null
160 Bloor Street East Ultrasou..	Toronto	15th Floor	160 Bloor Street East	7641890
360 Pharmacy	Richmond Hill	Unit S, 20-22	360 Highway 7 East	3590037

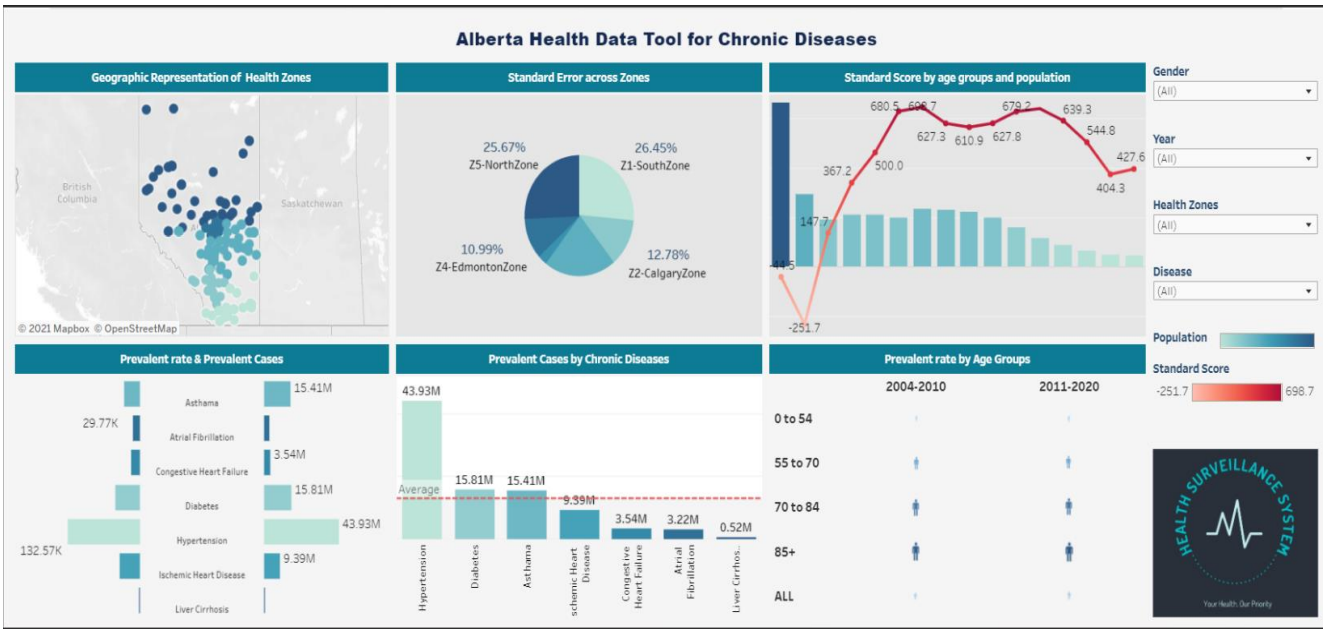
III Alberta Data Tool Dataset:



Sr No.	Name of Chronic Disease	Age-sex Standardized Prevalance
1	Asthama	
2	Atrial Fibrillation	
3	Congestive Heart Failure	
4	Diabetes	
5	Hypertension	
6	Ischemic Heart Disease	
7	Liver Cirrhosis	

Names	Geograpahy
Rainbow lake	Z5-NorthZone
High level	Z5-NorthZone
Fort vermilion	Z5-NorthZone
For Chipewyan	Z5-NorthZone
La Crete	Z5-NorthZone
Manning	Z5-NorthZone
Fort McKay	Z5-NorthZone

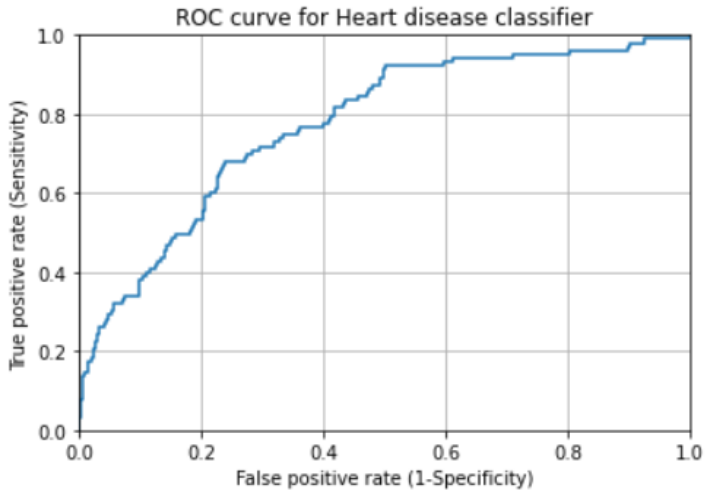
- Out of the chronic diseases selected, the cleaning was done and the combined csv were fed into the tableau software for visualization.
- The geography for the province of Alberta is subdivided into five health zones, hence to plot them on tableau we had to create a location wise csv and connect both the dataset together and obtain the geographic wise visual.
- The end visuals depict various insights filtered by various parameters that gives a detailed knowledge of these selected disease by location and gender.
- Furthermore, the visualizations were integrated into “Health Surveillance” API framework.

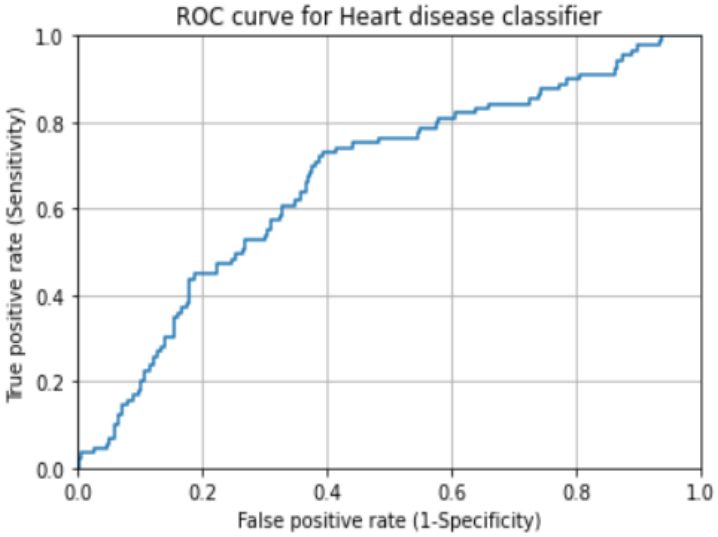


IV Disease Prediction Model:

- The dataset for the project contains 4238 observations. The target variable was 'TenYearCHD' which means coronary heart disease of a ten year span having binary values of '0' or '1'. Algorithm which yielded the highest AUC while analyzing the data in Jupyter notebook was Logistic Regression (AUC=0.77) whereas the AUC with Tabnet Neural Classifier was lower (AUC=0.66).

- The ROC curve for Logistic Regression:





- The feature selection technique used was Backward elimination to remove the attributes with the highest P-value one at a time followed by running the regression repeatedly until all attributes have P-values less than 0.05. Several factors like Odds Ratio, Confidence intervals and P-values were calculated to understand the increase or decrease of values of the variables for every unit increase.

Conclusion:

- The expected outcome from this project would be to deliver a seamless platform to respective hospitals that provides them to identify, control and monitor healthcare.
- The prediction model for heart disease shows a good predictive ability with a scope on incorporating and adding prediction model for other diseases.
- The project can be improvised by adding more modules for visualizations and by merging the location and supplier visualization into one as they belong to the same province.
- The projects offer and justifies the objective set at the initial stage and serves a platform to monitor various disciplines of the healthcare sector.

Contributions:

Name	Contributions
Alok Suresh Chilka	Create ML based web application for model using flask framework.
Pallavi Ravikumar Menon	Performed EDA and worked on all the tableau visualizations for all the datasets.
Sonia Rajput	Developed prediction model for heart disease.

References:

1. https://health-infobase.canada.ca/ccdss/data-tool/
2. https://prism.ucalgary.ca/handle/1880/106538
3. https://www.hindawi.com/journals/jhe/2020/4984967/
4. https://www.cdc.gov/heartdisease/journal_articles.htm
5. https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/types-of-heart-failure
6. https://www.webmd.com/heart-disease/guide-heart-failure
7. https://www.ahajournals.org/journal/jaha
8. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0248240
9. https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/types-of-heart-failure
10. https://www.kaggle.com/mansoordaku/ckdisease
11. https://health-infobase.canada.ca/ccdss/data-tool/Age?V=7&M=1&S=B&Y=2016