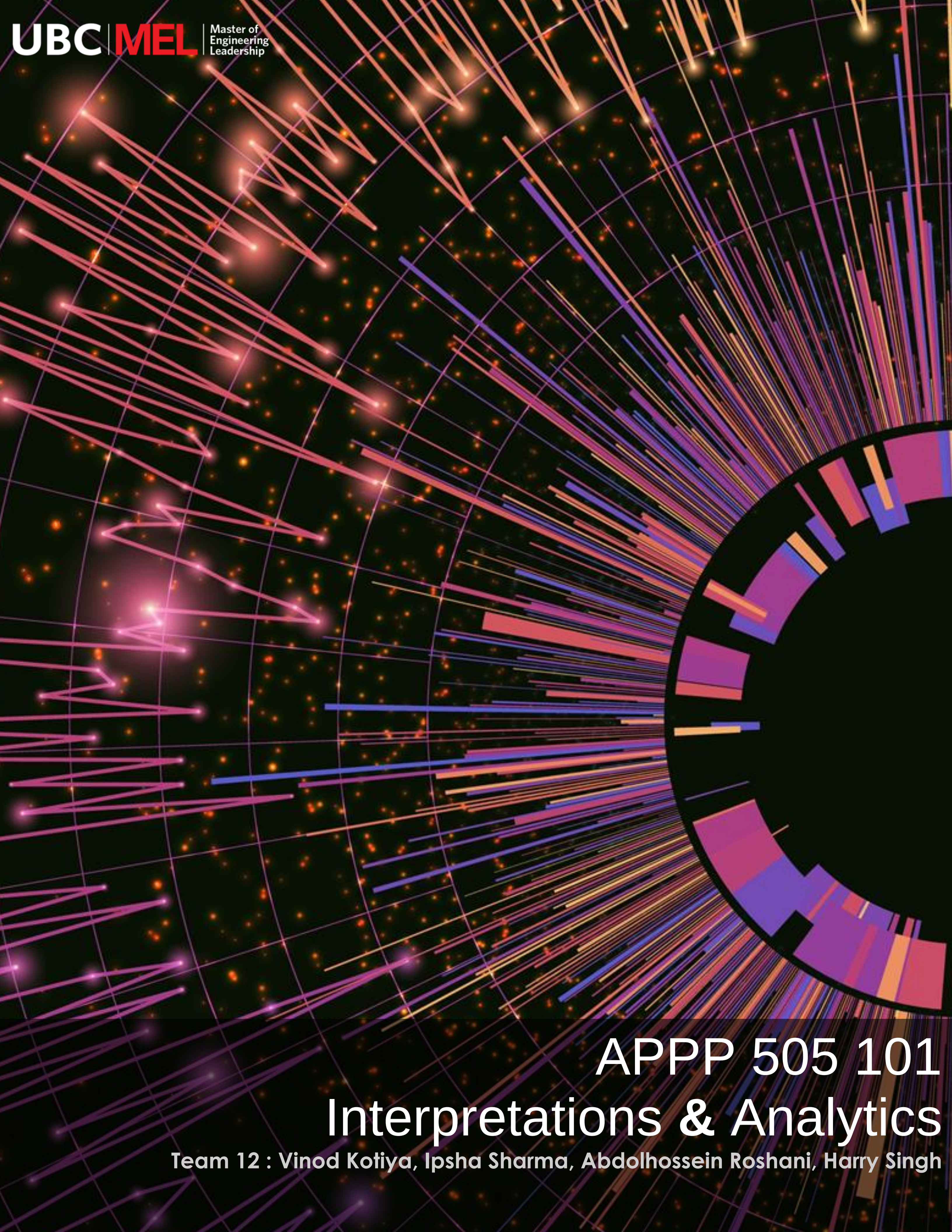
UBC | MEL Master of Engineering Leadership



APPP 505 101
Interpretations & Analytics

Team 12 : Vinod Kotiya, Ipsha Sharma, Abdolhossein Roshani, Harry Singh

Initial questions

The project focuses on the experience of recent immigrants living in Canada using the indicators like housing affordability and household expenditures collected and produced by Statistics Canada –

- What are the housing analysis and trend in different provinces of Canada?
- What are the categorical data of different housing ownerships?
- What are the housing percentage owned by male and female in highest populated provinces?
- How much Canadian households spend their money on main expenditure categories (Shelter, Transportation, Food, communication) in recent decade?
- What provinces are more affordable to live comparing with others in terms of those expenditure category.
- What is the trend of total expenditure in recent years?
- Prediction of housing expenditure in coming years.

Data preparation (Data Source: Two numbers of datasets from Statistics Canada's website under Open License)

- Canada Housing Data: Household spending, Canada, regions, and provinces. *Link (1.5 GB)*
- Canada Household Expenditure Data: *Link (6.5 MB)*
- Cleanup Process: Tableau used for Data analysis process which is available in *GIT Repository*.

Cleaning Housing Data:

In the housing data we identified irrelevant features for new immigrants were identified. Data was cleaned and further corrected for analysis. One of the major issues in analyses was multiple rows and high amount of data 1.5Gb. Following steps as per attached flow file in Tableau Prep where done.*(data flow file .tflx)*

- During the cleaning process we deleted many unnecessaries values like vector, scalar id, scalar factor, coordinate, symbol.
- We created a calculated field, where we identified the length of the province ID and asked tableau to consider only the last 2 digits of the DGUID field.
- We then created a clean step where we dumped all the false data and kept only the required data which matched the length of the matching province ID.
- Simultaneously we imported the meta-data with provincial information and created a clean step where we removed the unnecessary fields.
- We then created a calculated field where we cleaned the bracket for the provincial ID.
- Then we joined the Clean Big Data and Clean Meta Data to match the province ID.
- We created a clean step for the joined data to remove the duplicate fields.
- Then we created the output file to be exported to tableau visualisation.
- We used the Survey of Household Spending (SHS) data for our project.
- The main purpose of the (SHS) is to obtain detailed information about household spending, as well as limited information on dwelling characteristics and household equipment.
- The SHS primarily collects detailed information on household expenditures.
- It also collects information about the annual income of household members **(From personal income tax data),** demographic characteristics of the household, dwelling characteristics (**Eg. Type, age, & tenure)** and household equipment. **(Eg. Electronics and communications equipment)**

Cleaning Household Expenditure Data:

In raw data the expenditure ID detail was missing which we brought in after using the joining step with metadata.*(data flow file .tflx)*

- At first, we downloaded the meta file and main data file then we divided the metadata into two file to use the expenditure and region parent ID in our flow.
- In next step we created the Tableau prep flow and added those data to the flow. At first, we created three clean steps for each data file and removed the redundant fields as well as renaming the fields.

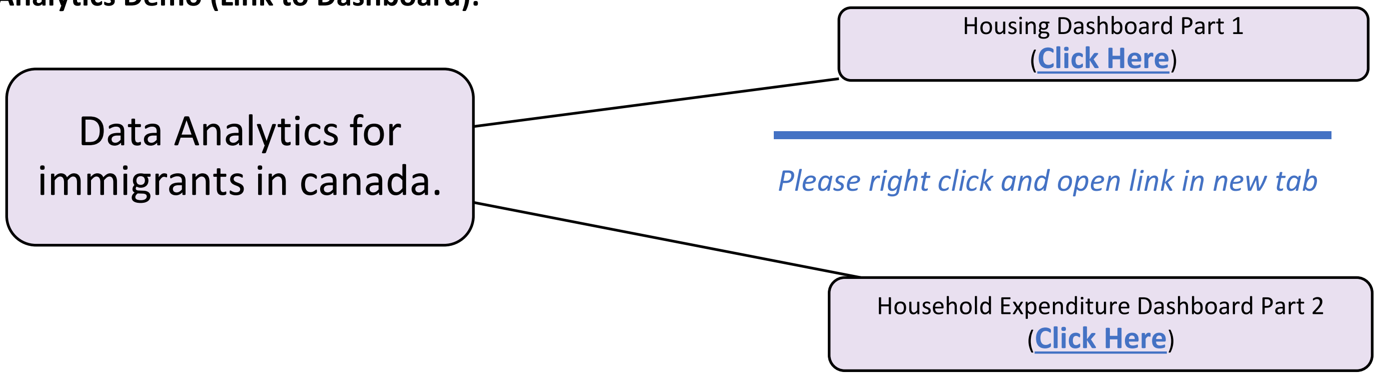
- After data cleaning we created two join steps to join the metadata files and main data files. Because we intended to add the parent ID field in meta file to the main dataset which helps us to determine the main category of expenditures as well as regions.
- Then we added another clean step to merge the duplicated fields.
- Finally, we created the output file to be exported to tableau visualization.
- We also created another cleaning step and output for our prediction by dropping some fields.

Explanatory Data Analysis (Using Tableau) -

We have made 2 dashboards for clearly explaining the housing scenario for a new immigrant in Canada. For Housing Dashboard Part 1 we have mostly used bar chart and crosstab for giving a comparative analysis of housing cast in different provinces according to family type. We opted for couple, and single family as major immigrants fall into this category, so we discarded the other features from stats Canada data source.

For Household Expenditure Dashboard Part 2 we have used bubble chart to show first impression of major annual expenditure to the immigrant. We have used a bar chart to give comparative analysis and a time series chart to show the increase in expenditure under different categories over the year. We again opted for features required for use of immigrants and discarded the irrelevant ones from stats Canada data source.

Analytics Demo (Link to Dashboard):



Data Analytics for immigrants in canada.

Housing Dashboard Part 1
(Click Here)

Please right click and open link in new tab

Household Expenditure Dashboard Part 2
(Click Here)

Findings: Based on above 2 Dataset we have created two dashboards for housing and household expenditure. Housing Dashboard: The housing analysis shows depiction of 3 major trends.

- **Chart 1: Couple Family owns housing around 36%-38% of total housing in all age groups -** We see that housing majority is concentrated at the top for couple family without children. With double income in one housing and less expenditure we can assume that is the reason for majority homes and with high average value homes for couple family. This trend is similar in all 3 age groups as under 35, 25-54 years old and above 55 years. This is followed by couple with children and then lone parents' category.

- **Chart 2: Average Housing in British Columbia is cheaper compared to Ontario & Employed Males own 3.6% more houses than females in Ontario -** We have compared 2 major provinces British Columbia and Ontario. Housing owned by Males and Females and further bifurcated into employed and unemployed category. Here we can see 2 major trends the value of housing in Ontario is much higher than British Columbia. Employed males own 3.6% more number of houses than employed females. But when we see the unemployed category females own 21.9% more houses than males. Whereas in British Columbia in both the employed and unemployed categories we see marginal difference between the housing owned by males and females.

- **Chart 3: Trend of Housing prices in 6 provinces from 2018-2020 -** We see a meteoric increase in housing prices for Ontario & British Columbia from 2018-2019.But when we see the other 4 provinces there is a steady increase. For New Brunswick and Nova Scotia there is a marginal increase in the consecutive years. In conclusion, we can say that Ontario and British Columbia get a larger influx of immigrants every year. This increases demand on the major cities, due to which we see the drastic rise in housing prices every year.

Household Expenditure Dashboard: This dashboard demonstrates a big picture of average Canadian households spending on main goods and services and its trends in recent years.

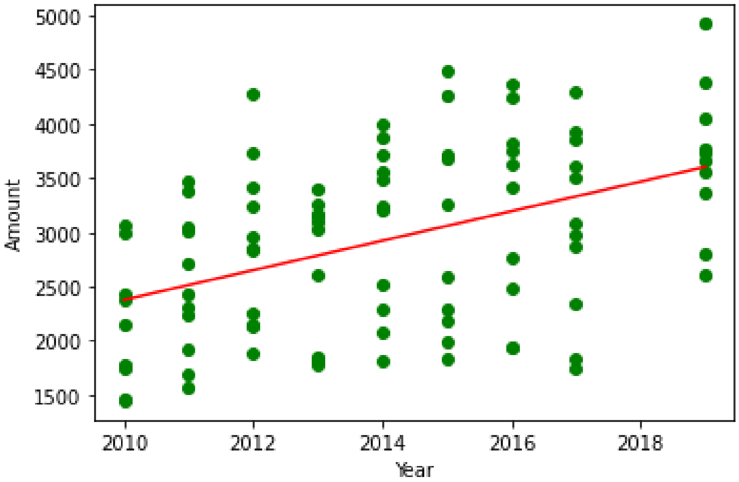
- **Chart 1: Main Average Expenditures Category Across Canada and Its Provinces -** This chart shows the categorized average expenditures across Canada and its provinces and regions in 2019. As can be seen in the chart, shelter has the largest portion of main expenses compared to others which is 20200 CAD in 2019 in Canada on average. By changing the filter (Check Box) we can see the data for any other provinces and territories.

- **Chart 2: Main Average Expenditure Category in Recent Decade -** This chart indicates the categorized average expenditures across Canada divided by recent years. As can be seen in the chart, shelter has the largest portion of main expenses compared to others which is 20200 CAD in 2019 in Canada on average. By changing the filter (Check Box) we can see the data for the year of interest.

- **Chart 3: Canadian Total Average Expenditure Trend in Last Decade (Provincewide Total Spending) -** This graph demonstrates the trends of total household spending in last decade across Canada and its provinces and territories, As indicated in the graph, total household spending expenditures experienced an increasing trend in recent years. It is worth noting that this trend is different from province to province. In 2019 Households in Alberta and British Columbia spent about (10000 CAD) more than average (68980 CAD) for their needs. In comparison Prince Edward Island is the affordable province in Canada with average (56662 CAD).

Predictive Data Analysis - In Progress (Using Python):

We have extracted the household expenditure data for doing the predictive analysis. We want to predict the expenditure in years 2019 onwards. We have already pre-processed the data for linear regression. After training the model using linear_model of sklearn_python package the scatter plot do not seem fit for prediction. We are getting error rate of Mean absolute error: 774.80, Residual sum of squares (MSE): 777790.56, and R2-score: -0.45. For this reason, we need more time to work upon polynomial regression or use some other prediction model.


Amount
Year