**Instructions for Power BI**

* Get the Data Visualization in [Data Visualization.pbix](https://mylambton-my.sharepoint.com/:u:/r/personal/c0866821_mylambton_ca/Documents/Documents/Lambton%20College/Class/Term%201/BDM-1043%20-%20Big%20Data%20Fundamentals/BDM-1043%20-%20Project/Instructions%20and%20Scripts/hadoop/Data%20Visualization.pbix?csf=1&web=1&e=p5OPvK)

|  |
| --- |
| Notes:   1. Auradee’s dataset will be used for the data visualization. **However, it is highly encouraged to have your own dataset in your local PC as well just in case Professor Jag will ask about it on presentation (see instruction on page 2).** 2. Do not update the tables on the Fields section, and the Power BI page that is not assigned to you 3. Do not change the Power BI canvas setting and background. Otherwise, inform the team first before the update. |

Graphical user interface, text, application, email

Description automatically generated

Power BI page assignment

* + **Age Group:** assigned to Bhumika and Pradeep
  + **NOC Skill Type:** assigned to Auradee
  + **Immigration Category:** assigned to Varun and Lakpa
* Color palette to use for uniformity. Go to this link <https://coolors.co/palette/8ecae6-219ebc-023047-ffb703-fb8500>, then click the color to copy the code.

Chart, treemap chart

Description automatically generated

When using the selected color on Power BI, simply paste the code in Hex field

Graphical user interface, PowerPoint

Description automatically generated

**Instructions on table creation and data loading in Hortonworks Hive.**

1. Transfer all [files for import](https://mylambton-my.sharepoint.com/:f:/r/personal/c0866821_mylambton_ca/Documents/Documents/Lambton%20College/Class/Term%201/BDM-1043%20-%20Big%20Data%20Fundamentals/BDM-1043%20-%20Project/Instructions%20and%20Scripts/hadoop/files%20for%20import?csf=1&web=1&e=UCW4L0) to Hortonworks local directory via WinSCP

Text

Description automatically generated

2. HIVE scripts for table creation, data loading and view creation. Execute the scripts sequentially.

Note: Change the path of the file before loading the data (highlighted in RED)

|  |
| --- |
| set hive.cli.print.current.db=true;  set hive.cli.print.header=true;  **-----------------------------------------**  **-- DATABASE CREATION**  **-----------------------------------------**  create database bdm1043;  use bdm1043;  **-----------------------------------------**  **-- PROVINCES**  **-----------------------------------------**  create table if not exists provinces (  province\_code string comment 'province code',  province\_name string comment 'province name',  region string comment 'region name'  )  row format delimited  fields terminated by ','  lines terminated by '\n'  stored as textfile  tblproperties ("skip.header.line.count"="1");  load data local inpath '/home/acastro/canada/provinces.csv'  into table provinces;  **-----------------------------------------**  **-- AGE\_GROUP**  **-----------------------------------------**  create table if not exists age\_group (  age\_group\_id int comment 'age group id',  age\_group\_desc string comment 'age group description'  )  row format delimited  fields terminated by ','  lines terminated by '\n'  stored as textfile  tblproperties ("skip.header.line.count"="1");  load data local inpath '/home/acastro/canada/age\_group.csv'  into table age\_group;  **-----------------------------------------**  **-- IMMIGRATION\_CATEGORY**  **-----------------------------------------**  create table if not exists immigration\_category (  immi\_code string comment 'immigration code',  immi\_type string comment 'immigration type'  )  row format delimited  fields terminated by ','  lines terminated by '\n'  stored as textfile;  load data local inpath '/home/acastro/canada/immigration\_category.csv'  into table immigration\_category;  **-----------------------------------------**  **-- NATIONAL\_OCCUPATIONAL\_CATEGORY**  **-----------------------------------------**  create table if not exists national\_occupational\_category (  noc\_code string comment 'national occupational category code',  job\_title string comment 'job title',  skill\_type string comment 'skill type: 0, A, B, C or D'  )  row format delimited  fields terminated by '|'  lines terminated by '\n'  stored as textfile  tblproperties ("skip.header.line.count"="1");  load data local inpath '/home/acastro/canada/national\_occupational\_category.csv'  into table national\_occupational\_category;  **-----------------------------------------**  **-- PR\_ADMISSION\_BY\_AGE\_GROUP**  **-----------------------------------------**  create table if not exists pr\_admission\_by\_age\_group (  age\_group\_id int comment 'age group id',  province\_code string comment 'province code',  year int comment 'year of the PR approval',  month string comment 'month of the PR approval',  total\_approval int comment 'total number of PR approvals'  )  row format delimited  fields terminated by ','  lines terminated by '\n'  stored as textfile  tblproperties ("skip.header.line.count"="1");  load data local inpath '/home/acastro/canada/pr\_admission\_by\_age\_group.csv'  into table pr\_admission\_by\_age\_group;  **-----------------------------------------**  **-- PR\_ADMISSION\_BY\_NOC**  **-----------------------------------------**  create table if not exists pr\_admission\_by\_noc (  noc\_code string comment 'national occupational category code',  province\_code string comment 'province code',  year int comment 'year of the PR approval',  month string comment 'month of the PR approval',  total\_approval int comment 'total number of PR approvals'  )  row format delimited  fields terminated by ','  lines terminated by '\n'  stored as textfile  tblproperties ("skip.header.line.count"="1");  load data local inpath '/home/acastro/canada/pr\_admission\_by\_noc.csv'  into table pr\_admission\_by\_noc;  **-----------------------------------------**  **-- PR\_ADMISSION\_BY\_IMMIGRATION\_CATEGORY**  **-----------------------------------------**  create table if not exists pr\_admission\_by\_immigration\_category (  immi\_code string comment 'immigration code',  province\_code string comment 'province code',  year int comment 'year of the PR approval',  month string comment 'month of the PR approval',  total\_approval int comment 'total number of PR approvals'  )  row format delimited  fields terminated by ','  lines terminated by '\n'  stored as textfile  tblproperties ("skip.header.line.count"="1");  load data local inpath '/home/acastro/canada/pr\_admission\_by\_immigration\_category.csv'  into table pr\_admission\_by\_immigration\_category;  **-----------------------------------------**  **-- VIEW FOR PR\_ADMISSION\_BY\_AGE\_GROUP**  **-----------------------------------------**  create view if not exists vw\_pr\_admission\_by\_age\_group  as select a.age\_group\_desc, p.province\_code, p.province\_name, pr.year, pr.month, pr.total\_approval  from pr\_admission\_by\_age\_group pr  inner join provinces p on pr.province\_code = p.province\_code  full outer join age\_group a on pr.age\_group\_id = a.age\_group\_id;  **-----------------------------------------**  **-- VIEW FOR PR\_ADMISSION\_BY\_NOC**  **-----------------------------------------**  create view if not exists vw\_pr\_admission\_by\_noc  as select n.noc\_code, n.job\_title, n.skill\_type, p.province\_name, pr.year, pr.month, pr.total\_approval  from pr\_admission\_by\_noc pr  inner join provinces p on pr.province\_code = p.province\_code  full outer join national\_occupational\_category n on pr.noc\_code = n.noc\_code;  **-------------------------------------------------**  **-- VIEW FOR PR\_ADMISSION\_BY\_IMMIGRATION\_CATEGORY**  **-------------------------------------------------**  create view if not exists vw\_pr\_admission\_by\_immigration\_category  as select i.immi\_type, p.province\_name, pr.year, pr.month, pr.total\_approval  from pr\_admission\_by\_immigration\_category pr  inner join provinces p on pr.province\_code = p.province\_code  full outer join immigration\_category i on pr.immi\_code = i.immi\_code; |

3. Steps to connect Power BI to Hortonworks

<https://chen115yaohua.wordpress.com/2017/01/27/microsoft-bi-access-hive-tables-on-hortonworks-sandbox/>

Choose ODBC when getting data from Hortonworks to Power BI

Graphical user interface, application

Description automatically generated