**Statistics Canada Data Exercise**

1. **Download and Ingest the following datasets in a data store of your choice (for more details on downloading the raw dataset, please see instructions at the end of this document).**
   1. Dataset 1: 98-400-X2016170 (<http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/Ap-eng.cfm?LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=0&GK=0&GRP=1&PID=110523&PRID=10&PTYPE=109445&S=0&SHOWALL=0&SUB=999&Temporal=2016,2017&THEME=119&VID=0&VNAMEE=&VNAMEF>=)
   2. Dataset 2: 98-400-X2016131 (<http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/Ap-eng.cfm?LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=0&GK=0&GRP=1&PID=110270&PRID=10&PTYPE=109445&S=0&SHOWALL=0&SUB=999&Temporal=2016,2017&THEME=119&VID=0&VNAMEE=&VNAMEF>=)
2. **Using Dataset 1, using any programming language, calculate and store the following metrics by Canadian regions based on the “Four-Region Model”** 
   1. Proportion of population by “Aboriginal Identity” and “Non-Aboriginal Identity”
   2. Average Total Income for “Aboriginal Identity” and “Non-Aboriginal Identity”
   3. Proportion of male vs. female population by “Aboriginal Identity” and “Non-Aboriginal Identity”
   4. Age group with most number of individuals with “Aboriginal identity”
      * For definition of “Four-Region Model” see following link (<https://en.wikipedia.org/wiki/List_of_regions_of_Canada>)
3. **Bonus 1: Using Dataset 1, create a heat map to visualize the total Aboriginal population by Canadian Geography**
4. **Bonus 2: Using Dataset 1, create an interface/API where users can visualize the population and Average Total Income by selecting the following Dimensions:**
   1. Canadian Region based on Four Region model
   2. “Aboriginal Identity” vs. “Non-Aboriginal Identity”
   3. Sex
   4. Age group
   5. City
5. **Bonus 3: Using any programming language, join Dataset 1 and 2 and create a single dataset that can provide the following metrics by Canadian province** 
   1. Population for “Aboriginal Identity”
   2. Population for “Non-Aboriginal Identity”
   3. Total population in 2015 in high income group (after-tax income >= $70,000) where an earner is present in the household

**Additional notes:**

* Feel free to use any tool you like to get the job done. You are allowed to use whatever solutions you choose to, provided you can explain the functions you are implementing in detail.
* Complete the assignment to the best of your ability with the time provided. Should you encounter challenges and inconsistencies with the problem or the data that prevents you from completing the objectives, please describe the challenges and/or trade-offs.

You may choose to host your work on GitHub, or email us directly.

**What are we looking for?**

* **We want to see how you handle:**

- Data transformation using programming languages

- Visualization tools to display output to end users

- Messy, real-world data

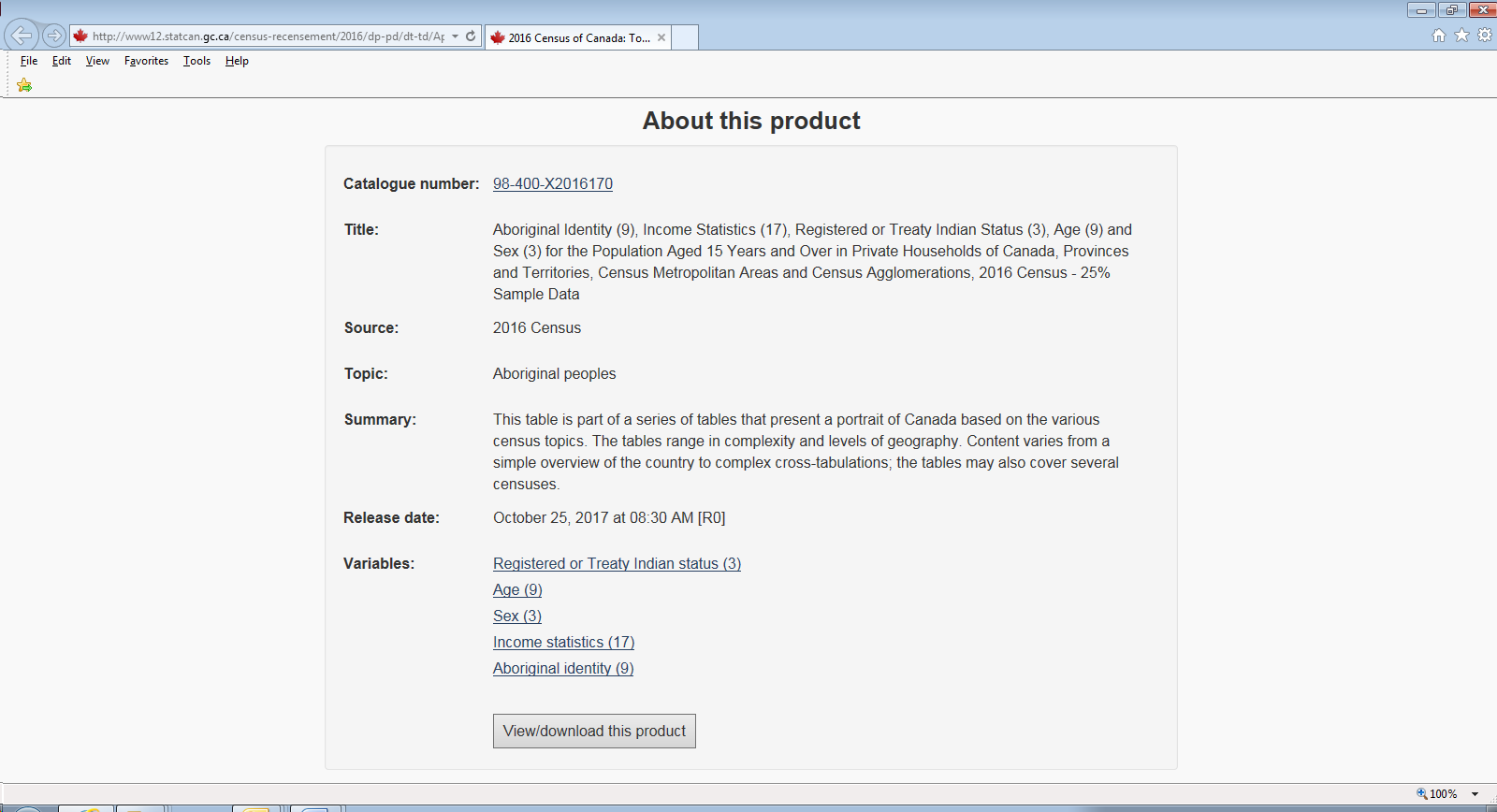
- Data transformation pipelines

- Algortihm design and analysis

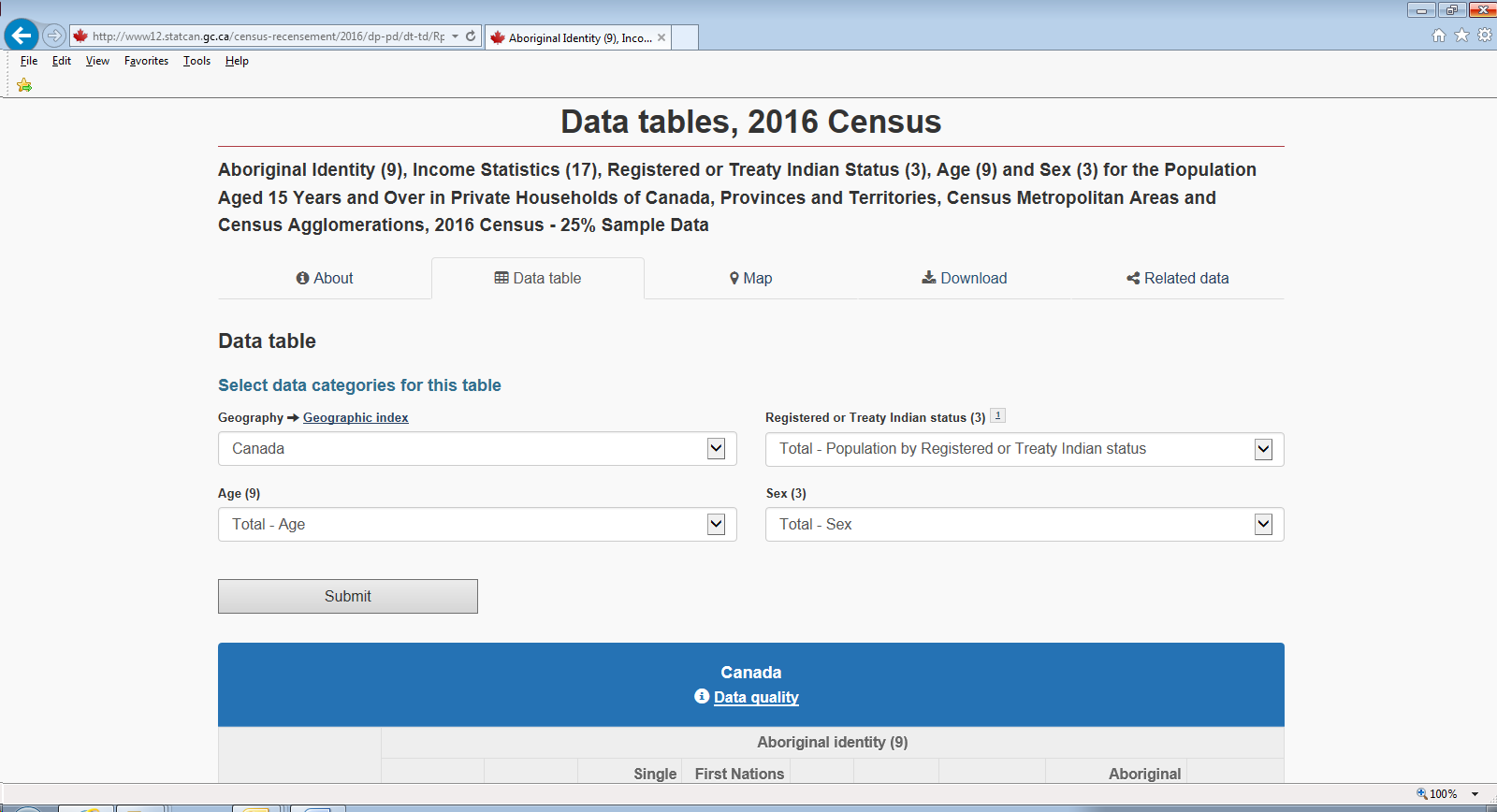
- Solving problems that involve working across the stack

**How to access the datasets?**

1. **Click on the link provided in the word document (see screen shot below)**



1. **Click on “View/Download this product” in the above screen shot**



1. **Click on “Download” in the above screen shot**
2. **Choose one of the “Download entire table” file formats in the screen shot below to access the raw datasets**

