INDIVIDUAL ASSIGNMENT 1 - DEFINING A DATA-DRIVEN BUSINESS CASE

- **Format: .doc or docx (max 2000 words)**
- **Grading: 20 points max (20% of final course grade)**

PART 1 – DATA EXPLORATION AND BUSINESS CASE

The goal of this first exercise is to explore the fundamentals of big data covered during the first class sessions and to apply the content to a potential data-driven business idea and existing (open) data sources. Sections:

- Definition of the business case: Context, problem statement, potential idea to create value with existing data sources
- Choice of a specific dataset (source, description, content, volume), format explanation (CSV, JSON, etc)
- Method of collection: API, direct download, etc.
- Exploratory analysis: type of data, useful fields, potential aggregations / modifications, etc.
- Data requirements: volume of information, free / paid access, combination with other existing or potential data sources, etc.
- Value proposition for potential clients or users: who would use it and what is in it for them
- Potential product definition: platform with dashboards, website, mobile app, etc.

PART 2 – THE DATA PRODUCT MANAGER ROLE

The goal of the second part is to understand how we organize data teams, helping you to reflect on the potential roles and responsibilities based on the business case explained in part 1. Sections:

- Definition of your project structure: roles, required resources (data, people, infrastructure, timeline), roles on your data team, etc.
- Creation of your product backlog: define the list of features you want to develop based on your previous analysis
- Specification of tasks to be done by your data team (i.e. analysis, development, etc.)
- Definition of questions or hypothesis to be sent to your data science team (i.e. number of women between 35 and 70 years old using a service during the weekend) so they can explore the data and get insights for you to use them as part of the product
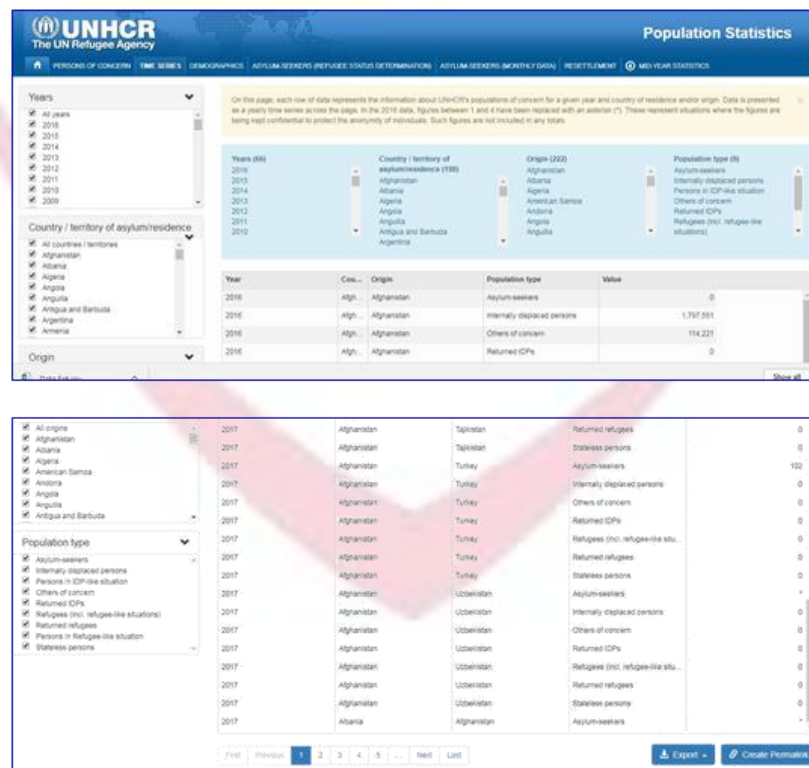
![Concordia University Continuing Education logo]

Big Data Program - CEBD 1150
Introduction to Big Data Business Cases & Data Collection
Instructor: Adrián González Sánchez
adrian.gonzalezsanchez@concordia.ca

INDIVIDUAL ASSIGNMENT 2 - DATA ANALYSIS & TRANSFORMATION WITH ALTERYX

PART 01 – ANALYZING AND TRANSFORMING DATA WITH ALTERYX

The objective of this exercise is to familiarize you with some of the challenges of exploring and analyzing data. We will use the data from the UNHCR (The UN Refugee Agency) and ask you to do work on the data to get some key information for a specific country (Syria).

INSTRUCTIONS

1. Please see the UNHCR site
2. Select all for:
   - Years
   - Country/ territory of asylum/residence
   - Origin
   - Population Type
3. Select the Export button (bottom of page) to get your file. This will be your input to Alteryx
4. For definitions of the fields, go to the overview page

QUESTIONS (NOTE: include detailed explanations, step-by-step analysis and screenshots)

1. Get and present some key information on refugees. This would need you to transform the data and organize it.
   a. Provide the top10 countries which have received the most refugees in a 2015
   b. Provide the top10 countries which are the highest "generators" of refugees in 2010
   c. Given the data you see, explain at least 5 other questions you answer with the data

2. Let's focus on The Syrian refugee crisis which has been in the news for a while now:
   a. Get the total population of Syria in 2010 and 2015 (from external websites)
   b. Which countries have hosted the most refugees from Syria (top 10) in 2015? How many?
   c. Provide a comparison between the number of refugees received by these countries (top 5) and their total population

3. Explain your insights into the Syrian refugee crisis from the data you have. What else of interest did you find in the data? How would you present the results?

4. What challenges did you face to get to your results? What areas are still not clear to you?

PART 02 – COMBINING DATA SOURCES FOR ADVANCED ANALYTICS

The objective of this exercise is to use Alteryx to transform and combine two or more (different) data sources. Optionally, you could try use Predictive Tools module of Alteryx to forecast results.

For this purpose, you can use either the source you chose for your first assignment or the one used for the first part of the second assignment (UNHRC), combine it with external data sources and analyze the output. You will have to explain carefully the rationale of your combination of data sources and the results obtained. You should add visualizations created via Alteryx or Excel.

INSTRUCTIONS FOR SUBMISSION

Please submit a .zip file containing:
- Your document answering the above questions (for both parts) in Microsoft Word format. The document should include screenshots of the tables or/and visualization you used. If you used external sources (web sites), please provide these sites
- Your Alteryx templates (one single .yxmd file per part)

EVALUATION

This assignment is 20% of the total grade (First part 12% + Second part 8%).
Both parts are mandatory and required to obtain the total grade.

TEAM PROJECT – REPORT AND PRESENTATION

PART 01 – TEAM REPORT: ANALYZING A NEW BIG DATA INITIATIVE

During the first part, you will analyze different business aspects of a Big Data project. For that purpose, you will need to choose an existing company and create a report for an internal Big Data initiative. Please keep in mind that you will need to understand the company's activities and to justify the value of Big Data technologies for this project. We will cover the content of this report during several sessions.

Format: Word/PDF file, 15 pages max (Calibri 12, Line Spacing 1'5)

Required sections:

1. Executive summary
2. Problem statement & choice of company
3. Definition of the big data business case
4. Business model canvas of the company
5. SWOT analysis of the company
6. ValueProp Canvas of the BigData initiative
7. Risk mitigation plan
8. Roadmap definition: Timeline, workstreams and milestones
9. Requirements: Infra, HR, software, etc.
10. Budget, ROI and soft benefits
11. Legal and ethical aspects

PART 02 – PRESENTATION: PITCHING A BIG DATA PROJECT

The second part of the project will be a presentation where you will be pitching your big data project to the "executives" of your company (roles played by external industry judges). All the team members are expected to present during similar amount of time.

Format: 15 min + 10 min (Q/A) – Number of slides is not limited, it will depend on your pitching style

Some aspects you may cover:

1. **Problem (or opportunity) statement:** Create the need for the company
2. **Introduction to potential solution:** Explain your vision
3. **Details of the solution:** Talk about data, infrastructure and human requirements
4. **Required investment:** Explain how much it will cost
   a. *VS Return On Investment:* Describe the hard (money) and soft benefits of your initiative
   b. *VS Value Proposition:* Explain why the company should invest in this project
5. **Strategic and technical roadmap:** Show what you will be doing and when it will happen
6. **Mitigation plan:** Use risks and challenges as a way to show that you have already a plan
7. **Sell yourself:** Present your credentials to your executives to show that you are THE team

NOTE: Every team member has to participate actively in both activities to get the team grade (up to 40%)