

Which topic did you choose to apply the data science methodology to? **(2 marks)**

credit cards in a scenario where we should use information of credit cards

Did the student pick one of the three topics proposed in the assignment overview?

- ☐ 0 points  
No
- ☒ **2 points**  
**Yes**



#### PROMPT

Next, you will play the role of the client and the data scientist.

Using the topic that you selected, complete the Business Understanding stage by coming up with a problem that you would like to solve and phrasing it in the form of a question that you will use data to answer. **(3 marks)**

You are required to:

1. Describe the problem, related to the topic you selected.
2. Phrase the problem as a question to be answered using data.

For example, using the food recipes use case discussed in the labs, the question that we defined was, "Can we automatically determine the cuisine of a given dish based on its ingredients?".

Finding business opportunities is a big challenge for entrepreneurs, the natural tendency is to start a related business close to one that is already successful, however there could be some missed opportunities for complementary products.

Evaluating the market situation and identify the potential growth of different kind of business from the ground could be very hard and with low accuracy, however credit cards has data of consumption per business that could be classified and sorted by several criteria, they also have information of each business, their products, location

#### RUBRIC

The student is required to come up a problem related to the topic they selected and the problem must be phrased as a question that can be answered using data. Use your best judgement to rate the student's completion of the Business Understanding stage.

- ☐ 1 point  
Poor. Some description is provided about the problem, but the question to be answered is missing.
- ☒ **2 points**  
**Good. The problem to be solved is described and a question is submitted but the question does not match the problem described.**
- ☐ 3 points  
Excellent. The student gave sufficient description of the problem, and the question to be answered reflects the problem described.



Briefly explain how you would complete each of the following stages for the problem that you described in the Business Understanding stage, so that you are ultimately able to answer the question that you came up with. **(5 marks):**

1. Analytic Approach
2. Data Requirements
3. Data Collection
4. Data Understanding and Preparation
5. Modeling and Evaluation

You can always refer to the labs as a reference with describing how you would complete each stage for your problem.

1. Analytic Approach As long as we are looking for business that some how could be related, it is necessary to use a descriptive model. It would be also useful defining clusters within certain area.

#### Data Requirements

Always keeping in mind our goal is to identify opportunities of related business, we would not need to collect the business names but its location (address), the type of business (food, cuisine, sports, etc). Historical data of consumption in each business. It will be necessary to collect information from several sources at least the most common used or accepted in most part of the business. It will also be necessary a list of products per industry type with sales information so the pre selected business could be expanded in related products. Credit cards use their information to provide re targeting services to other companies, detailed information of business is probably confidential, however consolidate information with out the names and fiscal registration may be available for commercial purposes. It would be necessary to contact the B2B department and request the information described above for, at least, the last 3 years so we could identify seasonality, and market trends. Products list could be find in any official national statistic organization. In Ecuador is the Central Bank the one who collects and publish data per product classified by the CIU code. This information should be one year late how ever, as long as there is enough data from several years it could be

The student is required complete each stage fo in the Business Unders judgement to rate the s stage.

- ☐ 1 point  
Poor. Many sta insufficient des
- ☐ 3 points  
Good. At least 1 and the descrip the question de Understanding stages are miss
- ☒ **5 points**  
**Excellent. All s appropriately clear and appl they defined i Understanding**

information shows high sales of some product and in the area there are few places where the item is sold and the second one when information of the area show a high sales of any product but complementary products aren't offered in the same area

#### Data Collection

Information from credit cards should be in a structured form but it would be necessary to obtain weekly and monthly consumption, as long as product sales information from Central Bank have also other data attached to each product, it would be necessary to remove that unwanted information. Data of products could be in a unstructured form so it would be necessary to define the fields, depending on the amount of rows it could be used Python or some other program in Jupiter notebooks.

**Data Understanding and Preparation** Some data could be incomplete so we should figure out if it is necessary or if we can skip it. Some other information could present the same product with different names so we would want to fix that. There will be in the product database a lot of information of different products such as gasoline that we will not be interested in so viewing the list of products and the amount of records should help us to concentrate in those we want to analyze. Then we should remove unnecessary data.

#### Modeling and Evaluation

As long as we're working on a descriptive model we would not need a training set, however getting a decent figures of the calculated Up To Date product consumption is necessary because we depend on this information to take the right decisions We should think that we're assuming that people who are consuming some product in a defined areas are likely to consume information to take the right decisions We should think that we're assuming that people who are consuming some product in a defined areas are likely to consume some related product, so the modeling process should take us to estimate the sales of a different suggested related products in a defined area. Some other market considerations may come out when performing the analysis so the process may be redefined and adjusted to get the best information to make a decision.