**Introduction to CRISP-DM.**

CRISP-DM, which stands for Cross-Industry Standard Process for Data Mining, is an industry-proven way to guide your data mining efforts. CRISP-DM is an iterative data mining mode and is a comprehensive methodology for data mining projects which provides a structured approach to guide data-driven decision making.

As a data methodology, a study of the CRISP-DM model includes six data mining stages, their descriptions and provides explanations of the relationships between tasks and stages. And as a process model, CRISP-DM provides high-level insights into the data mining cycle. Like other data mining science methodologies, CRISP-DM requires flexibility at each stage, and communication with peers, management, and stakeholders to keep the project on track. After any of the following six stages, data scientists might need to revisit an earlier stage and make changes.

**Business Understanding stage**

The business understanding stage is the most important because this stage sets and outlines the intentions of the data analysis project. This stage is common to both John Rollins data science methodology, and CRISP-DM methodology. This stage requires communication and clarity to overcome stakeholders ‘differing objectives, biases, and information related modalities. Without a clear concise and complete understanding of the business problem and project goals, the project effort will waste time and resources.

**Data Understanding Stage**

Then, CRISP-DM combines the stages of data requirements, data collection, and data understanding from Johns Rollins methodology outline into a single data understanding stage. During this stage, data scientists decide on data sources and acquire data.

**Data Preparation Stage**

Next during the data preparation stage, data scientists transform the collected data into a usable data subset and determine if they need more data. With data collection complete, data scientists select a dataset and address questionable missing or ambiguous data values. Data preparation is common to foundational data methodology in CRISP-DM.

**Modelling Stage**

The modelling stage fulfils the purpose of data mining and creates data models that reveal patterns and structures within the data. These patterns and structures provide knowledge and

insights that address the stated business problem and goals. Data scientists select models based on subsets of the data and adjust the models as needed. Model selection is an art and science. Both foundational methodology and CRISP-DM focus on creating knowledge information that has meaning and utility.

**Evaluation Stage**

During the evaluation stage, data scientists test the selected model. Data scientists usually prepare a pre-selected test to run the trained model. The test platform sees the data as new and data scientists then assess the model's effectiveness. These testing results determine the model's efficacy and foreshadow the model's role in the next and final stage.

**Deployment Stage**

Finally, during the deployment stage, data scientists and stakeholders use the model on new data outside of the scope of the dataset. New interactions during this stage might reveal the new variables and need for a different dataset and model. Remember that the CRISP-DM model is iterative and cyclical, deployment results might initiate revisions to the business needs and

actions, the model and data, or any combination of these items.

**Feedback Stage**

After completing all six stages, you'll have another business understanding meeting with the stakeholders to discuss the results. In CRISP-DM, the stage is not named. However, in John Rollins Data Science methodology model, the stage is explicitly named the Feedback stage. You'll continue the CRISP-DM process stages until the stakeholders, management, and you agree that the data model and its analysis provide the stakeholder with the answers they need to resolve their business problems and attain their business goals.

The CRISP-DM model consolidates the steps outlined in foundational data methodology into the following six stages, business understanding, data understanding, data preparation, modelling, evaluation, and deployment. You'll continue the CRISP-DM process until the stakeholders, management, and you agree that the data model and its analysis answer the business questions.

During this final project you'll complete **3** tasks for a total of **10** points to demonstrate your knowledge of CRISP-DM data methodology.

First, you'll take on ***both*** the role of the client ***and*** the data scientist to develop a business problem related to one of the following topics:

* E-mails
* Hospitals
* Credit Cards

You'll use the business problem you defined to demonstrate your knowledge of the Business Understanding stage.

Then, taking on the role of a data scientist, you'll describe how you would apply data science methodology practices at each of the listed stages to address the business problem you identified.

You'll enter your answers in the text fields provided online. After you submit your assignment, one of your peers who are is completing this assignment within the same session will grade your final project. You will also grade a peer's assignment.

Please note that this assignment is worth 10% of your final grade.

Note: You can take as many breaks as needed between the exercises.