Module 1 Lab 2: Test Your Knowledge

CS 82B

BRENNA THOMAS

1. What are the top three data science programming languages, in order, according to the 2020 Kaggle data science and machine learning survey?

***PYTHON, SQL, R***

1. What is the trade-off between using a GUI versus using a programming language for data science?

***GUI’s lack flexibility to customize techniques and results, but takes less work than a programming language.***

1. What are some of the GUIs for data science that we mentioned?

***EXCEL, SAS, RapidMiner, Alteryx, Tableau, QlikView***

1. What are the top three cloud providers for data science and machine learning according to the Kaggle 2020 survey?

***AWS, GCP, Azure***

1. What percentage of time do data scientists spend cleaning and preparing data? What specializations in and around data science did we discuss?

***25% - 75% of time is spent. Specializations: machine learning, BI, Deep Learning, Data Engineering, Big Data, Statistical Methods.***

1. What data science project management strategies did we discuss, and which one is the most recent? What are their acronyms and what do the acronyms stand for?

***CRISP DM – Cross-Industry Standard Process for Data Mining***

***TDSP – Team Data Science Process (MOST RECENT)***

1. What are the steps in the two data science project management strategies we discussed? Try to draw the diagrams of the strategies from memory.

***CRISP DM***

*Business Understanding Data Understanding Data Preparation*

*Deployment Evaluation Modeling*

***Microsoft TDSP***

*START BUSINESS UNDERSTANDING*

*MODELING*

*DATA ACQUISITION & UNDERSTANDING*

*DEPLOYMENT CUSTOMER ACCEPTANCE*

1. Summarize the skills required for Machine Learning, Business Intelligence and Data Engineering.

***Machine Learning: typical data science techniques like munging, but with extensive knowledge of machine learning methods with the ability to deploy machine learning models at scale.***

***Business Intelligence: Utilize data science techniques, then communicate business decision that should be made. BI specialists use GUI tools and code with Python or SQL.***

***Data Engineering: Data engineering encompasses skills that are often used in the first steps of the data science process, like collecting, organizing, cleaning, and storing. They have Linux skills and can deploy machine learning models.***