Data Science Assignment
Final Data Science Project

Assignment Title:	**Final Data Science Project**
Objective:	The objective of this assignment is to combine all of the skills and knowledge gained in this course and complete a data science project from beginning to end.
Instructions:	You have just received a letter from your University's Research Experience for Students department that says, "Congratulations! You have been accepted into our Data Science Research Experience Project. We have 3 research projects that you get to choose from. For each of the projects the goal is to use data science to build a model to help predict future outputs. The Student Mental Health Project has collected data about students and if they have decided to seeked treatment. The Student Entrepreneur Project has collected data about students and if they have decided to start a business. The final project is a business partnership with a telecom company to build a model to predict if a customer is going to cancel their service. We look forward to hearing back what project you choose to move forward with." For this assignment you will be able to choose between final_project_telcom_churn_dataset.csv, final_project_student_mental_health_dataset.csv, and final_project_student_entrepreneur_dataset.csv. The goal of the final project will be to build a model to predict the outcome (target feature) for the dataset that you choose. You will need to work through the data science process including data exploration, data visualization, data curation, feature engineering, and modeling training. **Data Exploration and Data Visualization (Module 4)** ● Take the time to complete data exploration on the data including finding standard deviation and variance for numerical features.

    - Remember this should be a continuous feature.
- Complete all data visualizations on the data to learn about the features.
    - Review distribution, skewness, and shape of the features.

Data Curation (Module 3)

- Identify any outliers, determine how you will deal with it and write what method you used for the outliers.
- Make sure to handle all missing data.
- Dummy Code all categorical we talked about for the features.

Feature Engineering (Module 7)

- Complete any transformations that need to be completed on the data.
    - Remember that you need to examine the data to determine which column needs a log transformation.
- Complete dummy coding for any categorical data that needs to be completed.
- Optional: Complete Principal Component Analysis on the dataset

Machine Learning Model (Module 5, Module 8, Module 9)

- Build and train a supervised learning model. This can be any supervised learning model including an ensemble model or deep learning model.
    - Remember to review the use cases for supervised learning and unsupervised learning to select the right method.
- Describe the evaluation of your model.
    - Need to include the accuracy, precision and Area Under the ROC Curve.

For the final assignment there will be 2 items that will need to be submitted into the folder with the link that will be provided to you. You will need to include your Jupyter Notebook that contains all of your code and comments. The

	other item you need to include is a presentation that explains your process. There should be at least 1 slide for each section of the assignment: Data Exploration and Data Visualization, Data Curation, Feature Engineering, and Machine Learning Model. The ideal presentation will explain why you chose the project and data set, your data preparation process, feature engineering choices, and what machine learning model you choose, why you choose it, and explanation of the results of the model. Student Presentation Examples: https://drive.google.com/drive/folders/1nWXB-eg-KZFKNjWxocnKon3RirNfHOnW
Resources:	Data Curation Resources: • https://www.mastersindatascience.org/learning/how-to-deal-with-missing-data/ • https://www.dataversity.net/what-is-data-curation/ • https://www.youtube.com/watch?v=DiKwYKmQzJc Data Visualization Resources • Matplotlib - https://matplotlib.org/3.5.0/ • Matplotlib Tutorials - https://matplotlib.org/stable/tutorials/index • Seaborn - https://seaborn.pydata.org/ • Seaborn Tutorials https://seaborn.pydata.org/tutorial.html Feature Engineering Resources • https://towardsdatascience.com/what-is-feature-engineering-importance-tools-and-techniques-for-machine-learning-2080b0269f10#:~:text=Feature%20engineering%20is%20the%20process,design%20and%20train%20better%20features. • https://towardsdatascience.com/feature-engineering-for-machine-learning-3a5e293a5114 Supervised Learning Resources • https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590 • https://towardsdatascience.com/beginner-friendly-resources-for-machine-learning-fd198f844dc3 • https://www.kaggle.com/ Deep Learning Resources • https://iq.opengenus.org/applications-of-rnn/ • https://theappsolutions.com/blog/development/convolutional-neural-networks/ • https://medium.com/datathings/the-magic-of-lstm-neural-networks-6775e8b540cd


AMERICA
ON TECH