**Daily Challenge Week 7 Day 4**

* Write a brief summary explaining the basics of Machine Learning and why it is important for data analysts.

There are various types of Machine learning such as Supervised, unsupervised and reinforcement learning. Generally it is used to predict a certain outcome based on previous information. The algorithms are capable of detecting patterns which are often difficult for a human to identify, and based on that make accurate forecastings. Data analysts often need to make data driven decisions and in order to do that, basic machine learning models could sometimes be a good way to make such get to conclusions. One example could be in marketing analysis. When a product is supposed to be marketed in different ways to different customer groups for ideal targeting and efficiant marketing strategy. to find the groups that have certain features in common to make them receptable to the same marketing campaign, ML could be used.

* Discuss the applications of Machine Learning across different industries, providing at least three specific examples.

Fraud Prevention: Identify if the purchase is with good intentions or fraudulant. Using data like previous purchase history, browser, device, location, etc. can increase/decrease likelyhood of a purchase beeing fraudulant.

Insurance: Should an applicant receive insurance? What is the likelyhood of them retreiving money from the insurance? Check for past claimes made, age, family history if available (and if legal), living area, occupation.

Gaming: clustering users into groups that are more likely to make purchases related to the game. Or, what features lead daily active users to increase their playing time.

* Describe the process of developing a machine learning model. Focus on three main stages: Feature Selection, Model Selection, and Model Evaluation.

Start by defining the problem and what we want to achieve with the model. Define a necessary accuracy of the model in order for it to be good enough for deployment. Collect enough relevant data. Clean that data by removing duplicates, replacing missing values, drop irrelevant columns. Change categorical values to dummy data if necessary. Change binary strings to numerical values. Make sure the data is balanced. Conduct normalization and/or standardization where necessary. Drop the column used as the target/label from the df and create a new variable for it. Call it y. the remaining features to be used will be x. use train test split from sklearn to split the data into training and testing datasets. Usually testing will be arounf 20-30% of the data. This can be dependant on the number of rows you have available. Choose a model that best encompasses your need. Are the features in linear correlation to the dependant variable? Are you looking for a binary classification? Or are you looking to cluster the variables? Train the model with the training dataset, then predict values with the test dataset. Based on those results, evaluate the models performance using metrics like confusion matrix, accuracy score, precision, recall, f1. If model did not perform well enough, change hyper parameters of the model and retrain it for better results. Continue this cycle until desired result is achieved.