Question 3

I would impute the mean value of that column for all null values. This will allow for no data loss and relatively accurate data. Discarding these rows is another option but that results in data loss, and the more data the better. However, for large amounts of missing values, dropping the rows might be the better option, I will explore which path to take later.

The tools I would use to do this are the pandas and sci-kit learn libraries in Python. Using the isnull() and sum() methods of pandas, I would first see how many values are missing in the column. If an outstanding percentage of the column is null I would drop those rows, however, I am going to assume that this is not the case. Leveraging the pandas fillna() method, I would impute null values with the mean value of the column. Additionally, imputing the most frequently occurring value is another technique I could use and can also be done using the fillna() method paired with the value_counts() method (index 0 of value counts will be the most common occurring). Moreover, sci-kit learn SimpleImputer allows for this functionality as well and would be my preferred choice. SimpleImputer takes the strategy you want to use and values you want to impute, in this case, the mean and null values. After fitting and transforming the column in question, the null values will be imputed with the mean of the column.

I would not use SQL since the pandas library allows for simpler and more concise syntax for imputing mean values. SQL is faster when dealing with larger file sizes like this one, but the pandas library allows for similar, if not faster, querying and writing speeds. Further, using the sci-kit learn library makes this process extremely simple and concise.

Question 4

I would automate this task using cron, which allows for the execution of a program at a specific time and day. The process would consist of creating a bash script of my program, scheduling the cron job via the command line and creating a cron expression consisting of the day and time (Thursday at 2 am) and bash script, and finally inserting the previously created bash script of my Python program.

Question 5

After watching *The Imitation Game*, I became fascinated with Alan Turing and his work towards modern computer science and mathematics. I was fascinated by Bombe and the Turing machine, which he created to break Enigma during World War II, and his implementation of algorithms, conditionals, and loops to decipher Enigma. Turing's work helped solve a trivial problem the world was facing during the height of World War II, and he did so through his genius, dedication, and persistence. Similarly, I want my work to amount to something greater and create real change in the world, whether it be by increasing sales, positively affecting consumers and consumer satisfaction, or launching a business forward through predicted and actionable business insights. I have a passion for helping others and creating positive change through my work, and I am eager to start applying my skills to real-world problems.