1. **Write-Up on ML Models. (Point: 2)**

Linear regression is among the simplest machine learning models, plotting a line to predict where similar data points would appear in a dataset. However, the model also expects data to be linear. Temperature data, however, will fluctuate as the seasons change, creating a nonlinear wave over years. This model is unlikely to be the correct choice.

Random forest instead uses a collection of decision trees, each trained on random subsets of the data. These trees vote on each data point and classify it with a majority, intending to create a set of classifiers that can eliminate errors through quantity. The randomness of the subsets are intended to vary their errors, which could make a good fit for this project.

Gradient boosting also uses a collection of models, but for a different purpose: Each model is sequentially trained, using weight updates based on the errors of previously misclassified instances. This causes each model to “fix” the last one’s error, continuing until all points are correctly classified (or a maximum epoch count is reached). Outliers could greatly mess with this algorithm because of this, so it may not be the best choice due to unusually hot or cold days.

With the above information in mind, a random forest classifier will likely be the best choice.

Retail Question: Customer Lifetime Value (CLV)

* + - How can we predict long-term revenue potential to prioritize high-value customers?

1. **Web Server Setup. (Point: 2)**

* Launch and configure a web server in Azure (or another platform, as long as it's internet-accessible).

A screenshot of a computer screen

Description automatically generated

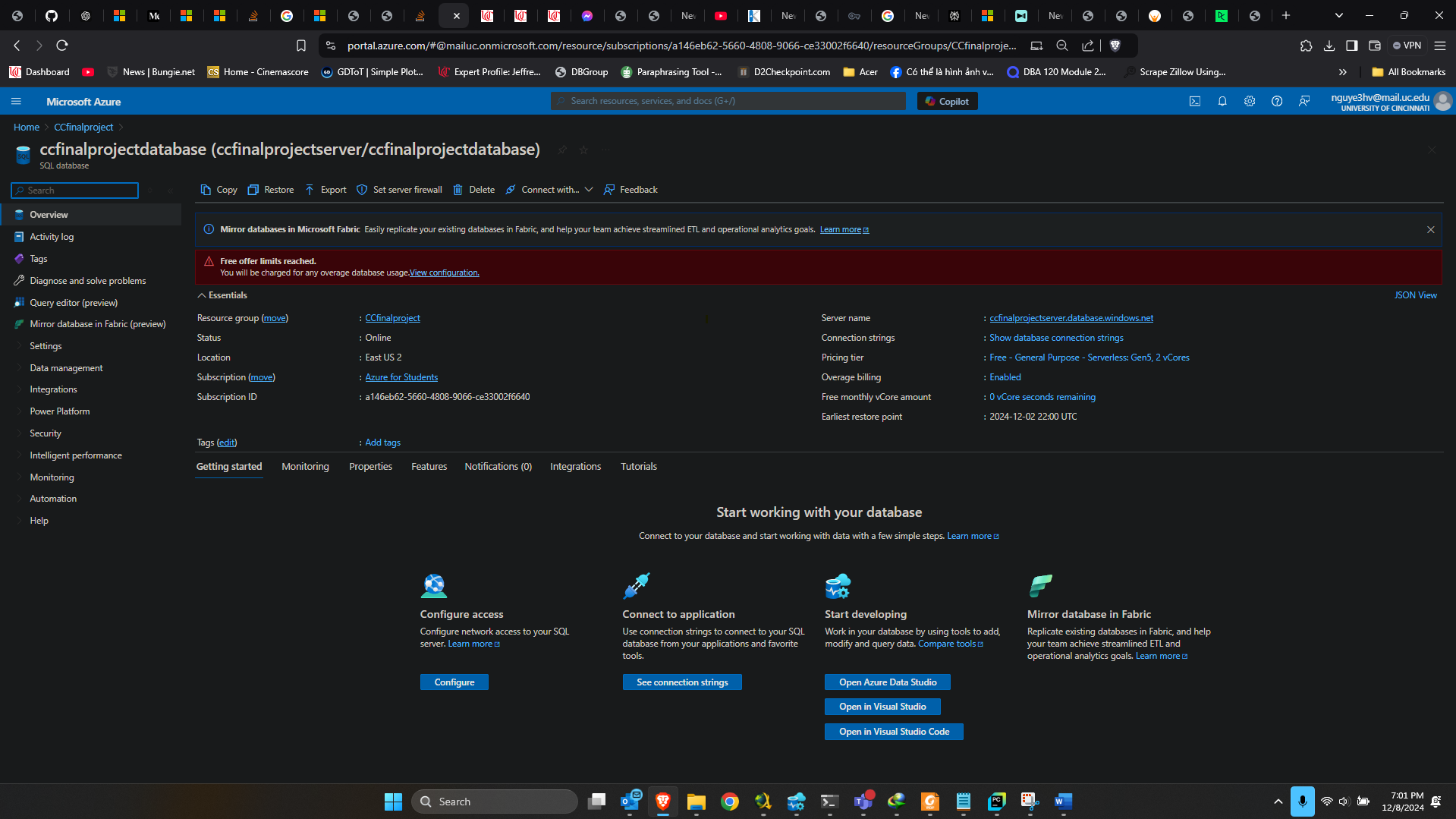
* Design an interactive webpage with the following fields:
  + Username
  + Password
  + Email

A screenshot of a computer

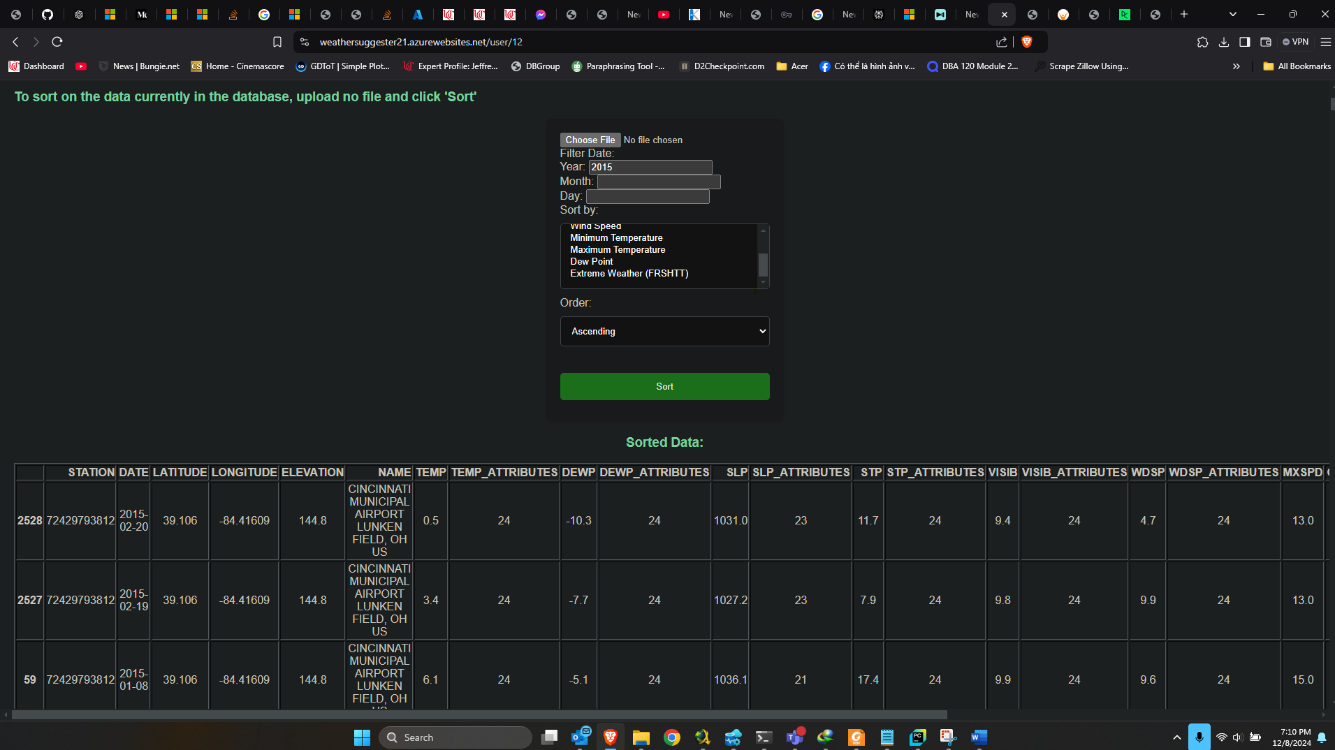
Description automatically generated

1. **Datastore and Data Loading. (Point: 2)**

Create a datastore or database in Azure (e.g., Azure SQL, PostgreSQL, MySQL, MongoDB, Azure Storage Account) and load sample data



Create a display page for a Sample Data Pull for HSHD\_NUM #10, linking the Households, Transactions, and Products tables. Sort by Hshd\_num, Basket\_num, Date, Product\_num, Department, Commodity.  Similar to the SAMPLE DATA PULL FOR HH #0001 table shown above.



1. **Interactive Web Page. (Point: 2)**

Create a webpage that allows users to search for Data Pulls based on Hshd\_num. Sort results by Hshd\_num, Basket\_num, Date, Product\_num, Department, Commodity.

A screenshot of a computer

Description automatically generated

1. **Data Loading Web App. (Point: 3)**

Create a web app that allows loading of the latest Transactions, Households, and Products datasets. Test the output on the interactive web page from Requirement #4 to ensure it functions with updated data.

1. **Web Page with Dashboard. (Point: 3)**

Design a webpage with a dashboard to explore retail challenges using selected factors from the provided "Examples of Questions to Address." Creativity is encouraged.

1. **ML Model Application. (Point: 2)**

Use one of the following ML models—i. Linear Regression, ii. Random Forest, iii. Gradient Boosting?

1. **Churn Prediction. (Point: 2)**