Week 5 Progress Report
DATA 450 Capstone

Bijay Adhikari

Week ending: 2/23/25

Time Log

Day	Time	# Hours	Task(s)	GH Commit(s)
Mon.				
Tue.	[10am-12pm; 7-12pm]	[7]	[Data collection from Steam API; Implemented retry strategy and rate limiting]	Link 1
Wed.	[4-6pm]	[2]	[Data cleaning; Removed invalid dates; Handled missing values]	Link 2
Thu.	[6pm-9pm]	[3]	[Exploratory data analysis; Created visualizations for active and inactive games]	Link 3
Fri.	[1pm-4pm]	[3]	[Analyzed playtime distribution and user engagement trends for Q1]	Link 4
Sat.				
Sun.				

Total # of hours worked this week: [15]

What you accomplished (or attempted) this week:

This week, I collected data from the Steam API , which involved generating random Steam User IDs and handling large-scale data requests. I implemented a retry strategy and used rate limiting with time.sleep() to avoid getting blocked by Steam's API. After data collection, I cleaned the dataset by removing invalid dates and handling missing values . I performed exploratory data analysis , created visualizations , and worked on Q1.

Any setbacks/roadblocks you experienced:

I faced challenges during the data collection process due to rate limiting by the Steam API , even though the limit is 100,000 requests per day . The server would stop accepting requests around 25,000 calls , leading to the need for manual retries and adjusting request intervals using time.sleep() . Handling HTTP request errors and server blocks also required additional error handling and retry strategies.

What you plan to work on next:

I plan to work on Question 2 . I will also gather more data to expand the dataset; since I already have the code to exact the data, it's only matter of changing the dataset number(how many data to look for) in the python script.

Are you on track? If not, how will you get back on track?

So far, yes , I am on track.