#-----------------------------From PPT:------------------------------------

Before we demonstrate our record, let’s review our data flow briefly. In our data flow, we first store our dataset in S3, then we use SQL to query in Athena. After that, we use Python to analyze and do the visualization, and in the end, we host a static website to show our results.

#----------------------------In Amazon S3 :----------------------------------------

Let’s first move to the AWS console and open Amazon S3.

For S3, we use the bucket that we have created in Assignment 4, Storage with CLI. For here, we use bucket edu.gwu.ds.cc.g35236263. Click this bucket name. In this bucket, we created a new folder named Final VG Sales to distinguish it from other documents. Click the Final VG Sales folder here. In this Final VG Sales folder, we upload our project dataset, video-game-sales csv, which is 1.3 MB.

#------------------------------------In Athena SQL:------------------------------------

Secondly, we use SQL query in Athena, to answer four questions that we have previously.

We again move to the AWS console, find Athena, click this. We create a new database called cc final, in this database, we create a new table named vg sales. Click run here to generate the table.

After we generated the table, we started to query those 4 questions.

For question 1, we use SQL to show the first 50 rows. Click run here. It shows the top 50 video games with the most global sales and their information.

For question 2, we query the released year before 2000 among the top 50 games, and show global sales in descending order. Click run bottom. Here are video games in the top 50 that were released before 2000.

For question 3, we group by the publisher, find the total North American sale and order it by descending order to find the top 10 popular publishers. Here are the top 10 highest sales publishers in North America.

Then we filter the top 10 publishers in Europe, Japan, and Other regions with the highest sales. This is for Europe, this is for Japan, and this is for other regions.

For the last question, we group by genre, find the total global sales and order it in descending order to find the genre with the highest global sales.

After we find the query result table, we can download and save each in a CSV file on the local folder.

#---------------------------------------In Jupyter Notebook:------------------------------------------

Then we can analyze our results in Python. We open a Jupyter notebook in anaconda and create a new Jupyter notebook named VG Sales Python Visualization.

Here is the python code we wrote in our jupyter notebook. We will take a quick look here because we will interpret detailly later in our static website.

After we finish the Jupyter notebook, we will save an html file to easily host a static website in AWS. Click file, click download as html and save as local folder.

#----------------------------------------In AWS Static Website:-----------------------------------------

In the end, we host a static website to shadow our html file. Go back to the AWS console, click S3. Click the bucket we used in the previous step, choose Properties, under Static website hosting, choose Edit, choose Enable.

We enter index document name: VG Sales Python Visualization.html, enter error document: error.html, choose Save changes

Then, choose permissions, under block public access, choose Edit, clear Block all public access, and choose Save changes. Edit bucket policy like it.

In our bucket, we click the properties again and go down to the bucket website endpoint. Click the link and open in a separate browser window.

#-----------------------------------------Analysis the graph:------------------------------------------------

Here is the static website. Below is the output of the expected outcome that we mentioned before. To better interpret for the audience, we use plotly to create interactive data analysis.

For question 1, we want to know the top 50 games which currently have higher global sales. We used pandas to read the csv file that we downloaded from Athena. The result table shows the top 50 games with high sales globally. Take a look at this table, Wii Sports has the highest global sales with 82.74 million, which is doubled for the second-highest global sales Super Mario Bros.

Question 2 is: are there any games in the top 50 with releases years older than 2000 that are still making high global sales? What are they? To answer this question, we read the csv file that we saved for question 2 in Athena. The result table shows the top 50 video games released years older than 2000. Based on these query results, Super Mario Bros has the highest global sales and its released year is 1985.

Question 3 is: what is the most popular publisher in each regional sale?

Let's look at North America first. The bar chart shows the most popular publisher in North America is Nintendo, which has 815.44 million. The rest of the popular publishers are Electronic Arts, Activision, Sony Computer Entertainment, Ubisoft, Take-Two Interactive, THQ, Microsoft Game Studios, Atari, and Sega.

For Europe, we repeat the same steps as we did in North America and show the top 10 most popular publishers in Europe sales. From the bar chart we created, the most popular publisher in Europe is Nintendo too with 418.52 millions. Besides, the first 7 most popular publishers in Europe are the same as in North America.

Similarly in Europe, we repeated the previous step in Japan and showed the top 10 most popular publishers. In Japan, Nintendo has the highest sales, 453.35 million, which is much more popular than other publishers. What’s more, we also can see that the sale of Nintendo is 4 times higher than Namco Bandai games.

For other regions, same as what we did previously, the green bar chart shows the most popular publisher in electronic arts with 129.77 million.

Question 4 is: which type of game is the highest in terms of global sales? To answer this question, we read the CSV file that we saved for question 4 in Athena, and create a pie chart to show each genre with global sales. The pie chart below shows that action video games have the highest global sales with 1744.3 million. The rest of the higher global sales are sports, shooter, role-playing, platform, misc, racing, fighting, simulation, puzzle, adventure, and strategy.

All right, that is all we want to present in our demo. Thanks for listening.

-by Pondeepan Raj