Lab 2 - Data Analytics Workflow

## Instructions

1. Answer the below question in the boxes provided.
2. Use the provided Video Game Sales dataset from this Chapter to complete your analysis in Part 2.
3. Please submit the assignment through TalentLabs Learning System.

## Part 1: Testing your knowledge of the data analytics workflow

Question 1.1:

What **three** iterative steps does Data Wrangling involve?

|  |
| --- |
| 1. Exploration  2. Transformation  3. Validation |

Question 1.2:

What data type would car brands fall under?

|  |
| --- |
| **Categorical** |

Question 1.3:

Give **two** reasons why it is important to document the analysis process?

|  |
| --- |
| 1. Future reference  2. Ensure process repeatable |

Question 1.4:

Name **one** reason why spreadsheets are a good tool for data analysis?

|  |
| --- |
| Easy to learn |

Question 1.5:

What information about stakeholders should you gather at the beginning of the analysis process? Give **two** examples.

|  |
| --- |
| 1. What are the expectations of the stakeholders?  2. Who are the stakeholders of the project? |

## Part 2: Video Game Sales Analysis

Task: Your employer was very pleased with your previous analysis and hired you again to answer five more questions on the same top 100 Video Game Sales dataset.

Question 2.1:

Name the newest and oldest game in the dataset.

|  |
| --- |
| Newest: Star Wars Battlefront  Oldest: Pac-Man |

Question 2.2:

How many sales (in millions) did the Puzzle Genre generate **in Europe**? (Hint: Use a Pivot Table)

|  |
| --- |
| 8.07 million |

Question 2.3:

How many total sales (in millions) did Mario Kart 7 generate? (Hint: Use the sort or search functionality)

|  |
| --- |
| 12.21 million |

Question 2.4:

How many games developed on the SNES platform made it to this top 100 list? (Hint: Use conditional formatting)

|  |
| --- |
| 4 games |

Question 2.5:

What is the **average** (in millions) of **North American** sales for games that are in the **Racing Genre** and are on the **PS3 Platform**? (Hint: Use Pivot Tables with multiple rows)

Your answer should be 1 number, for example: 5.43 million

|  |
| --- |
| 2.96 million |