LSE DA301: Advanced Analytics - Turtle Games Project

This repository contains the coursework for the LSE Data Analytics Career Accelerator course DA301. The project involves a comprehensive analysis of the Turtle Games dataset to derive business insights on customer loyalty, segmentation, and sentiment. The analysis is conducted using both Python for machine learning tasks and R for statistical modeling and visualization.

Repository Structure

The project is organized into separate directories for data, Python notebooks, R scripts, and final reports. The R environment is managed by `renv` for full reproducibility.

```
LSE_DA301_BP/
├── Data/
│   ├── metadata_turtle_games.txt
│   ├── turtle_reviews.csv
│   └── turtle_reviews_2.csv
├── Python/
│   ├── .venv
│   ├── Berni_Alberto_DA301_Assignment_Notebook.ipynb
│   └── requirements.txt
├── R/
│   ├── Berni_Alberto_DA301_Assignment_Notebook.Rmd
│   └── LSE_DA301_Assignment_R_template.R
├── Reports/
├── renv/
│   ├── activate.R
│   └── settings.json
├── .Rprofile
├── .gitignore
├── renv.lock
└── README.md
```

Environment Setup

To ensure full reproducibility of the analysis, please follow the setup instructions below.

R Environment

The R analysis uses the `renv` package to create a reproducible, project-specific environment.

1. **Prerequisites**: Before you begin, ensure you have the necessary system-level software installed.

R: You must have a base installation of R. You can download it from the Comprehensive R Archive Network (CRAN).

Pandoc: The `rmarkdown` package, which is used to generate reports, requires **Pandoc** as a system dependency. RStudio typically bundles this automatically, but for standalone R setups (like in VS Code), it must be installed separately.

Download: Go to the official Pandoc installation page.

- **Windows**: Download and run the `.msi` installer. The installer will automatically add Pandoc to your system's PATH, which is required.
- **macOS**: The recommended installation method is using Homebrew. In your terminal, run:

```
brew install pandoc
```

- **Linux**: Use your distribution's package manager. For example, on Debian/Ubuntu, run:

```
sudo apt-get install pandoc
```

Verify Installation: After installing, close and reopen your terminal or IDE, then run:

```
pandoc --version
```

You should see the installed version number (`1.12.3 or higher`).

2. **Clone the Repository**: If you haven't already, clone the project to your local machine.

```
git clone <your-repository-url>
cd LSE_DA301_BP
```

3. **Open the Project in R**: Start an R session with the project's root directory as the working directory. The easiest way to do this is by opening the project in an IDE like RStudio or VS Code (with the R extension).

4. **Restore the `renv` Environment**: When you first open the project, the `.Rprofile` script should automatically detect the lockfile. If prompted, agree to restore the environment from the lockfile. If you are not prompted, simply run the following command in the R console:

```
renv::restore()
```

This command reads the `renv.lock` file and installs the exact versions of all required R packages, ensuring a fully reproducible setup.
You are now ready to run the analysis in `R/Berni_Alberto_DA301_Assignment_Notebook.Rmd`.

Python Environment

The Python analysis is managed within a dedicated virtual environment.

1. **Clone the repository** to your local machine:

```
git clone <your-repository-url>
cd LSE_DA301_BP
```

2. **Navigate to the Python directory**:

```
cd Python
```

3. **Create a virtual environment**:

```
python -m venv .venv
```

4. **Activate the environment**:
   - On **macOS/Linux**:

```
source .venv/bin/activate
```

   - On **Windows**:

```
.\.venv\Scripts\activate
```

5. **Install the required dependencies**: With the environment activated, install all necessary libraries from the `requirements.txt` file.

```
pip install -r requirements.txt
```

6. **Register the environment as a Jupyter Kernel**: This step makes your virtual environment available within Jupyter, ensuring your notebook uses the correct packages.

```
python -m ipykernel install --user --name=lse-da301-project
```

7. **Launch Jupyter and Run the Notebook**:
    - From terminal:

      ```
      jupyter notebook
      ```

    - From your IDE (VS Code):
      Open `Berni_Alberto_DA301_Assignment_Notebook.ipynb` from your IDE.
      Go to `Kernels > Change Kernel`
      Select `Jupyter Kernel` and select **`lse-da301-project`**.