**ReadMe file**

**CSV folder**

1. Detail Correlation matrix – **corerelation\_matrix.xlsx**
2. Output data file with session – **output\_session.csv**
3. Output with epoch to timestamp conversion – **output.csv**

**Python & R code Folder**

1. Univariate & Bivariate analysis code - **Warmup.py**
2. Join two file (User & end Songs) and convert epoch to timestamp -**Data\_merge\_epoch\_conversion.py**
3. Session Design and put session number in raw data also Translation of variables - **Session\_design\_variable\_transformation.py**
4. Session and ms\_played related analytics & correlation matrix define – **Analytics.py**
5. K-means algorithm perform in R – **K-means.R**

**PPT**

1. Deck about the findings from this whole analytics & data engineering work. - **sportify\_ppt.pptx**
2. Command use for AWS/Linux/S3/EMR & **External table** definition in AWS EMR - **Sportify Analysis commands**

**Images**

1. Cluster visual representation (it’s very bad to interpret) - **Rplot\_cluster.png** & **Rplot01\_cluster.png**