Apple Store Project

July 16, 2019

In [7]:
```
# This program reads the data from the Apple store file
# Source: https://www.dataquest.io/blog/basic-data-science-portfolio-project-tutorial/
```

In [18]:
```
import os
os.getcwd()
```

Out[18]:
```
'/Users/johnfields/Desktop/MyPythonFiles/Project'
```

In [27]:
```
### The App Store data set ###
from csv import reader
opened_file = open('applestore.csv')
read_file = reader(opened_file)
ios = list(read_file)
ios_header = ios[0]
ios = ios[1:]
```

In [29]:
```
def explore_data(dataset, start, end, rows_and_columns=False):
    dataset_slice = dataset[start:end]
    for row in dataset_slice:
        print(row)
        print('\n') # adds a new (empty) line between rows

    if rows_and_columns:
        print('Number of rows:', len(dataset))
        print('Number of columns:', len(dataset[0]))
```

In [30]:
```
print(ios_header)
print('\n')
explore_data(ios, 0, 3, True)
```

```
['id', 'track_name', 'size_bytes', 'currency', 'price', 'rating_count_tot', 'rating_count_ver'


['281656475', 'PAC-MAN Premium', '100788224', 'USD', '3.99', '21292', '26', '4', '4.5', '6.3.5


['281796108', 'Evernote - stay organized', '158578688', 'USD', '0', '161065', '26', '4', '3.5'
```

```
['281940292', 'WeatherBug - Local Weather, Radar, Maps, Alerts', '100524032', 'USD', '0', '1885
```

```
Number of rows: 11100
Number of columns: 17
```

```
In [64]: ### Frequency tables with percentages ###
         def freq_table(dataset, index):
             table = {}
             total = 0

             for row in dataset:
                 total += 1
                 value = row[index]
                 if value in table:
                     table[value] += 1
                 else:
                     table[value] = 1

             table_percentages = {}
             for key in table:
                 percentage = (table[key] / total) * 100
                 table_percentages[key] = percentage

             return table_percentages
```

```
In [69]: ### Display percentages in a descending order ###
         def display_table(dataset, index):
             table = freq_table(dataset, index)
             table_display = []
             for key in table:
                 key_val_as_tuple = (table[key], key)
                 table_display.append(key_val_as_tuple)

             table_sorted = sorted(table_display, reverse = True)
             for entry in table_sorted:
                 print(entry[1], ':', entry[0])
```

```
In [70]:
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)

<ipython-input-70-f95ac009fb94> in <module>
----> 1 for entry in table_sorted:
```

```
      2         print(entry[1], ':', entry[0])


    NameError: name 'table_sorted' is not defined
```

```
In [37]: ios_final = []
```

```
In [38]: def is_English(string):
             non_ascii = 0

             for character in string:
                 if ord(character) > 127:
                     non_ascii += 1

             if non_ascii > 3:
                 return False
             else:
                 return True
```

```
In [39]: ios_english = []
```

```
In [56]: ### ios ###
         for app in ios:
             name = app[1]
             if is_English(name):
                 ios_english.append(app)

         ### Check number of entries left ###
         explore_data(ios_english, 0, 3, True)
```

```
['281656475', 'PAC-MAN Premium', '100788224', 'USD', '3.99', '21292', '26', '4', '4.5', '6.3.5
```

```
['281796108', 'Evernote - stay organized', '158578688', 'USD', '0', '161065', '26', '4', '3.5'
```

```
['281940292', 'WeatherBug - Local Weather, Radar, Maps, Alerts', '100524032', 'USD', '0', '188
```

```
Number of rows: 20172
Number of columns: 17
```

```
In [63]: print(ios[1:10][1])
```

```
['281940292', 'WeatherBug - Local Weather, Radar, Maps, Alerts', '100524032', 'USD', '0', '188
```

In [66]:
```
### Generating a frequency table to get the unique app genres ###
genres_ios = freq_table(ios_final, -5)

### Looping over the unique genres ###
for genre in genres_ios:
    total = 0
    len_genre = 0

    ## Looping over the App Store data set ##
    for app in ios_final:
        genre_app = app[-5]
        if genre_app == genre:
            n_ratings = float(app[5])
            total += n_ratings
            len_genre += 1

    ## Compute and display the average number of user ratings ##
    avg_n_ratings = total / len_genre
```

In [67]:

```
---------------------------------------------------------------------------

NameError                                 Traceback (most recent call last)

<ipython-input-67-8dcb74a128a8> in <module>
----> 1 print(genre, ':', avg_n_ratings)


NameError: name 'genre' is not defined
```

In [ ]: