Name:_____ Period:

Do Now: Data Review (5 minutes/ 5 points)

Try to describe the types of information you will find in each column of our 2019 Billboard data set using only one word (Text, String, Integer, Number, Boolean, Float, Date, etc.)

WeekID	Week Number	Week Position	Song	Performer	Previous Week Position	Peak Position	Weeks on Chart
1/5/19	1	37	A Lot	21 Savage		37	1
1/5/19	1	40	Blue Christmas	Elvis Presley		40	1
1/5/19	1	57	Swervin	A Boogie Wit da Hoodie Featuring 6ix9ine		57	1
1/5/19	1	58	Can't Leave Without It	21 Savage		58	1
1/5/19	1	65	Startender	A Boogie Wit da Hoodie Featuring Offset & Tyga		65	1

WeekID	Week Number	Week Position	Song	Performer	Previous Week Position	Peak Position	Weeks on Chart

Mini Lesson: Data Analysis with Python (10 minutes/ 5 points)

Python commands	Definitions
`import pandas as pd` `import plotly.graph_objects as go`	"Import" is a command that introduces a new set of tools into your environment. Think of your coding space as an empty garage, and each import adds a tool to your garage that wasn't there before (like a tablesaw, or an electric drill)
`df = pd.read_csv`	Df stands for "dataframe". pd.read_csv is telling our imported tool "pandas" to read our excel file (the csv) and to store the contents of the data in a variable called "df"

Once our excel data is inside our coding environment as the dataframe variable "df", we can use python to interact and manipulate with the data. We'll be using these commands

`df['Column Name']=="thing"`	Selects specified values from specific column
`df.sort_values(by=['Weeks on Chart'], ascending=False)`	Sorts table by specified column values
`df.drop_duplicates(['Column name'])`	Removes duplicate row entries in specified column
`smaller_df = df.head(10)`	Creates a new dataframe with the first 10 rows of the original dataframe

Activity: Building dataframes (20 minutes/ 10 points)

Given the prompts and starter code below, create and print dataframes from our "Billboard_Top_100.csv" that produce the values described in the prompt.

1. Create and print a dataframe of the top 10 most featured artists descending

```
artist_instances = df['Performer'].value_counts().rename_axis('artist_name').reset_index(name='counts')
top_10 = artist_instances.head(____)
print(top_10)
```

2. Create and print a dataframe of the top 10 songs spending the most weeks on the charts

```
sorted_df = df.sort_values(by=['____________________'], ascending=False)
no_duplicates = sorted_df.drop_duplicates(['Song'])
top_10 = no_duplicates.head(___)
print(top_10)
```

3. Pick one song that has been on the chart for at least 25 weeks and create a dataframe of its weekly position

```
song_condition = df['_______']=="_____"
song_df = df[___________]
print(song_df)
```

Exit Ticket

Create a hypothesis about what you think the code segment below does. Explain your hypothesis by referring to specific sections of the code

```
fig = go.Figure([go.Bar(x=top_25['artist_name'], y=top_25['counts'])])
fig.show()
```