Day 20: *Combining Our Knowledge*

After figuring out that Seattle has the best connection with Vancouver, Dot hopped on a flight heading north and east. The plane bounded over the endless Pacific until it began its descent over the Puget Sound, eventually landing in Washington State among the Space Needle and the tall office buildings where tech companies worked. What a different environment than Hawaii! Dot had some time to kill before they boarded their connection to Vancouver. Instead of wasting it in the airport, they checked out the city. Dot had never been to Seattle before and wanted to visit the Museum of Pop Culture for the first time.

The building rose up strangely and majestically before Dot's eyes, resembling a quilt of colourful, hallucinatory sheets of metal blowing in the wind. Dot had never seen anything like it. Inside, they strolled through the exhibits paying tribute to humanity's music, art, music, and video games. Dot was inspired by how much beauty and culture people had contributed through the years. All this reflecting on the past made Dot reflect on theirs – specifically, what they'd done over the past weeks. Dot had visited many different cultures across the earth, taking in the sights and sounds of cities in the Americas, Europe, Asia, and Australia. It had been incredible to see how different but similar societies across the world were. As they rode back to the airport to make their connecting flight, Dot smiled with happiness as they sifted through their memories. Will you join Dot in reminiscing as you review what you've learned over the past few weeks?

Tutorial

Through the past 19 days, we've gone over the fundamentals of Python, Pandas, SQL and Matplotlib. With these technologies combined, you're well on your way to doing great data science and analytics work. In today's challenge, we'll be integrating aspects from all of the previous challenges, combining them to do a data analysis of the world cities we've visited.

Challenge

Pull the data from table `worldcities` in the database world.db to Pandas data frame to answer the following questions:

1. Using histogram from matplotlib, visualize distribution of `population` across all cities. What do you observe?
2. Create bar chart with the population of top 10 the most populated countries. Which country is two most populated?

Stretch

- What is the largest city Dot visited in the last three weeks?
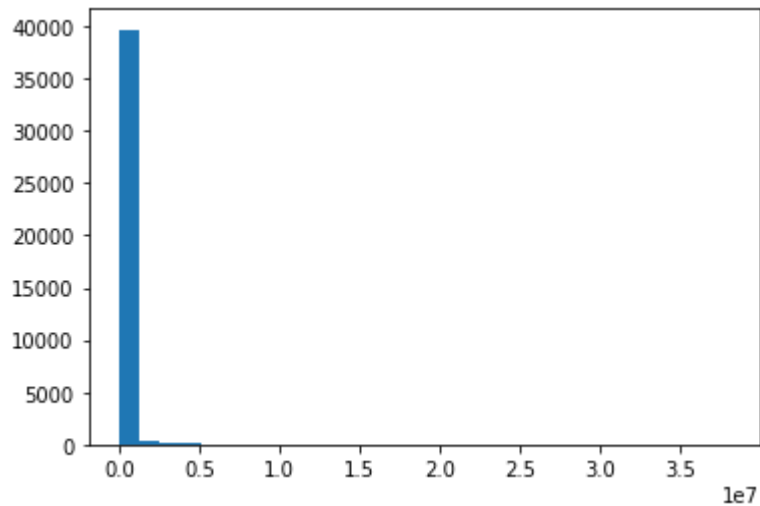- What is the difference in population between the largest and smallest one?

```
In [1]:  import pandas as pd
         import matplotlib.pyplot as plt
         import sqlite3
```

In [2]:
```python
conn = sqlite3.connect("world.db")
```

In [3]:
```python
# SOLUTION

df = pd.read_sql('select * from worldcities',conn)
```

In [4]:
```python
plt.figure()
plt.hist(df['population'], bins = 30) #Play around with the bin sizes when plotting you
plt.show()
```


In [5]:
```python
countries = df.groupby("country")['population'].sum().sort_values(ascending = False)
```

In [9]:
```python
countries.head(1).index.values[0]
```

Out[9]:
```
'China'
```