1. The Oldest Businesses in the World

Staffelter Hof Winery, Germany's oldest business, which was established in 862 under the Carolingian dynasty is said to be the oldest business in the world. It has continued to serve customers through dramatic changes in Europe such as the Holy Roman Empire, the Ottoman Empire, and both world wars. What characteristics enable a business to stand the test of time?

To help answer this question, BusinessFinancing.co.uk researched the oldest company that is still in business in almost every country and compiled the results into a dataset. Let's explore this work to to better understand these historic businesses. Our datasets, which are all located in the datasets directory, contain the following information:

1. A businesses.csv and new_businesses.csv dataset
2. A countries.csv dataset
3. A categories.csv dataset

In [5]:
```
# Import the pandas library with its usual alias
import pandas as pd

# Load the businesses.csv dataset as a DataFrame
businesses = pd.read_csv(r"C:\Users\USER\Downloads\project\What and Where are the World'

# Have a glimpse of the businesses DataFrame
businesses.head()
```

Out[5]:

	business	year_founded	category_code	country_code
0	Hamoud Boualem	1878	CAT11	DZA
1	Communauté Électrique du Bénin	1968	CAT10	BEN
2	Botswana Meat Commission	1965	CAT1	BWA
3	Air Burkina	1967	CAT2	BFA
4	Brarudi	1955	CAT9	BDI

In [13]:
```
# Sort the businesses from oldest to youngest
sorted_businesses = businesses.sort_values("year_founded", ascending=True)

# Display the first few rows of the sorted businesses
sorted_businesses.head()
```

Out[13]:

	business	year_founded	category_code	country_code
64	Kongō Gumi	578	CAT6	JPN
94	St. Peter Stifts Kulinarium	803	CAT4	AUT
107	Staffelter Hof Winery	862	CAT9	DEU
106	Monnaie de Paris	864	CAT12	FRA
103	The Royal Mint	886	CAT12	GBR

In [17]:
```
#An overview of the range of the year_founded
year_founded_range = sorted_businesses['year_founded'].agg(['min', 'max'])
```

```
#Display year_founded_range
year_founded_range
```

Out[17]:
```
min     578
max    1999
Name: year_founded, dtype: int64
```

This dataset shows the businesses between 578AD and 1999

2. The oldest businesses in North America

So far we've learned that Kongō Gumi is the world's oldest continuously operating business, beating out the second oldest business by well over 100 years! It's a little hard to read the country codes, though. Wouldn't it be nice if we had a list of country names to go along with the country codes? The countries.csv dataset is located in our dataset folder. Let's see the oldest business in North America.

In [4]:
```
# load the countries.csv dataset as a DataFrame
countries = pd.read_csv(r"C:\Users\USER\Downloads\project\What and Where are the World's

# Have a glimpse of the countries DataFrame
countries.head()
```

Out[4]:

	country_code	country	continent
0	AFG	Afghanistan	Asia
1	AGO	Angola	Africa
2	ALB	Albania	Europe
3	AND	Andorra	Europe
4	ARE	United Arab Emirates	Asia

In [54]:
```
# Merge the sorted_business with countries
businesses_countries = sorted_businesses.merge(countries, on="country_code")

# Filter business_countries to list only old businesses in North America
north_america = businesses_countries[businesses_countries["continent"] == "North America

# Display the first few rows of the north_america DataFrame
north_america.head()
```

Out[54]:

	business	year_founded	category_code	country_code	country	continent
22	La Casa de Moneda de México	1534	CAT12	MEX	Mexico	North America
28	Shirley Plantation	1638	CAT1	USA	United States	North America
33	Hudson's Bay Company	1670	CAT17	CAN	Canada	North America
35	Mount Gay Rum	1703	CAT9	BRB	Barbados	North America
40	Rose Hall	1770	CAT19	JAM	Jamaica	North America

3. The Oldest Businesses On Each Continent

Now we can see that the oldest company in North America is La Casa de Moneda de México in Mexico, founded in 1534. Why stop there, though, when we could easily find out the oldest business on every continent?

In [50]:
```
# Create DF, continent, which lists only continents and their oldest year_founded
continent = businesses_countries.groupby("continent").agg({"year_founded" : 'min'})
continent = continent.sort_values("year_founded", ascending=True)

# Display part of the DF continent
continent.head()
```

Out[50]:

continent	year_founded
Asia	578
Europe	803
North America	1534
South America	1565
Africa	1772

In [52]:
```
# Merge continent with business_countries
merged_continent = continent.merge(businesses_countries, on="year_founded")

# Display a few of the DF merger_continent
merged_continent
```

Out[52]:

	year_founded	business	category_code	country_code	country	continent
0	578	Kongō Gumi	CAT6	JPN	Japan	Asia
1	803	St. Peter Stifts Kulinarium	CAT4	AUT	Austria	Europe
2	1534	La Casa de Moneda de México	CAT12	MEX	Mexico	North America
3	1565	Casa Nacional de Moneda	CAT3	PER	Peru	South America
4	1772	Mauritius Post	CAT16	MUS	Mauritius	Africa
5	1809	Australia Post	CAT16	AUS	Australia	Oceania

In [20]:
```
# Subset continent so that only interested columns are included
subset_merged_continent = merged_continent[["business", "continent", "country", "year_fo

# Display a few content subset continent
subset_merged_continent.head()
```

Out[20]:

	business	continent	country	year_founded
0	Mauritius Post	Africa	Mauritius	1772
1	Kongō Gumi	Asia	Japan	578
2	St. Peter Stifts Kulinarium	Europe	Austria	803
3	La Casa de Moneda de México	North America	Mexico	1534
4	Australia Post	Oceania	Australia	1809

4. Unknown Old Businesses

BusinessFinancing.co.uk wasn't able to determine the oldest business for some countries, and those countries are simply left off of *businesses.csv* and, by extension, *businesses*. However, the *countries* that we created *does* include all countries in the world, regardless of whether the oldest business is known.

We can compare the two datasets in one DataFrame to find out which countries don't have a known oldest business!

In [21]:
```
# Create a DataFrame, all_countries, merging both businesses and countries
all_countries = countries.merge(businesses, on="country_code", how="outer")

# Display few content of the DF
all_countries.head()
```

Out[21]:

	country_code	country	continent	business	year_founded	category_code
0	AFG	Afghanistan	Asia	Spinzar Cotton Company	1930.0	CAT1
1	AGO	Angola	Africa	NaN	NaN	NaN
2	ALB	Albania	Europe	ALBtelecom	1912.0	CAT18
3	AND	Andorra	Europe	Andbank	1930.0	CAT3
4	ARE	United Arab Emirates	Asia	Liwa Chemicals	1939.0	CAT12

In [22]:
```
# Filter to include only countries without oldest_businesses
missing_countries = all_countries[pd.isna(all_countries["year_founded"])]

# Display few missing countries
missing_countries.head()
```

Out[22]:

	country_code	country	continent	business	year_founded	category_code
1	AGO	Angola	Africa	NaN	NaN	NaN
7	ATG	Antigua and Barbuda	North America	NaN	NaN	NaN
18	BHS	Bahamas	North America	NaN	NaN	NaN
48	DOM	Dominican Republic	North America	NaN	NaN	NaN
50	ECU	Ecuador	South America	NaN	NaN	NaN

In [41]:
```
# Create a series of the country names with missing oldest business data
missing_countries_series = missing_countries[['country', 'continent']]

# Display a few row of the DataFrame
missing_countries_series.head()
```

Out[41]:

	country	continent
1	Angola	Africa
7	Antigua and Barbuda	North America
18	Bahamas	North America
48	Dominican Republic	North America
50	Ecuador	South America

5. Adding New Old Businesses Data

It looks like we've got some holes in our dataset! Fortunately, we've taken it upon ourselves to improve upon BusinessFinancing.co.uk's work and find oldest businesses in a few of the missing countries. We've stored the newfound oldest businesses in _new*businesses*, located at "C:\Users\USER\Downloads\project\What and Where are the World's Oldest Businesses\datasets\new_businesses.csv". It has the exact same structure as our businesses dataset.

In [27]:
```
# load new_businesses.csv
new_businesses = pd.read_csv(r"C:\Users\USER\Downloads\project\What and Where are the Wo

# Display the DataFrame
new_businesses
```

Out[27]:

	business	year_founded	category_code	country_code
0	Fiji Times	1869	CAT13	FJI
1	J. Armando Bermúdez & Co.	1852	CAT9	DOM

In [29]:
```
# Add the data in new_businesses to the existing businesses
all_businesses = pd.concat([new_businesses, businesses])

# Create a  new DataFrame, new_all_countries, merging both all_businesses and countries
new_all_countries = all_businesses.merge(countries, on="country_code", how="outer")

# Filter to include only countries without oldest_businesses
new_missing_countries = new_all_countries[pd.isna(new_all_countries["year_founded"])]

# Display a few rows of the new_missing_countries
new_missing_countries.head()
```

Out[29]:

	business	year_founded	category_code	country_code	country	continent
165	NaN	NaN	NaN	AGO	Angola	Africa
166	NaN	NaN	NaN	ATG	Antigua and Barbuda	North America
167	NaN	NaN	NaN	BHS	Bahamas	North America
168	NaN	NaN	NaN	ECU	Ecuador	South America
169	NaN	NaN	NaN	FSM	Micronesia, Federated States of	Oceania

In [42]:
```
# Group by continent and country and create a "count_missing" column
count_missing = new_missing_countries.groupby(['continent', 'country']).agg({'country':'
count_missing.columns = ['missing_count']

# Display a few rows of the counts
count_missing.head()
```

Out[42]:

		missing_count
continent	**country**	
Africa	**Angola**	1
	Gambia	1
	Ghana	1

Asia	**Iran, Islamic Republic of**	1
	Kyrgyzstan	1

In [32]:
```
# Group by continent and count total missing businesses by continent
count_missing = count_missing.reset_index()
count_missing_total = count_missing.groupby('continent').agg({'country':'count'})
count_missing_total.columns = ['missing_count_per_continent']

# Display the counts
count_missing_total
```

Out[32]:

continent	missing_count_per_continent
Africa	3
Asia	7
Europe	2
North America	5
Oceania	10
South America	3

6. The Oldest Industries

Remember our oldest business in the world, Kongō Gumi? We know Kongō Gumi was founded in the year 578 in Japan, but it's a little hard to decipher which industry it's in. Information about what the _category*code* column refers to is in "C:\Users\USER\Downloads\project\What and Where are the World's Oldest Businesses\datasets\categories.csv".

Let's use categories.csv to understand how many oldest businesses are in each category of industry.

In [34]:
```
# Import categories.csv
categories = pd.read_csv(r"C:\Users\USER\Downloads\project\What and Where are the World'

# Display a few rows
categories.head()
```

Out[34]:

	category_code	category
0	CAT1	Agriculture
1	CAT2	Aviation & Transport
2	CAT3	Banking & Finance
3	CAT4	Cafés, Restaurants & Bars
4	CAT5	Conglomerate

In [36]:
```
# Merge categories with all_businesses
business_categories = categories.merge(all_businesses, on="category_code")
business_categories = business_categories.sort_values("year_founded", ascending=True)

# Display a few rows
business_categories.head()
```

Out[36]:

	category_code	category	business	year_founded	country_code
71	CAT6	Construction	Kongō Gumi	578	JPN
65	CAT4	Cafés, Restaurants & Bars	St. Peter Stifts Kulinarium	803	AUT
89	CAT9	Distillers, Vintners, & Breweries	Staffelter Hof Winery	862	DEU
119	CAT12	Manufacturing & Production	Monnaie de Paris	864	FRA
118	CAT12	Manufacturing & Production	The Royal Mint	886	GBR

In [40]:
```
# Create a DataFrame which lists the number of oldest businesses in each category
count_business_categories = business_categories.groupby("category").agg({'business':'cou
count_business_categories = count_business_categories.sort_values('business', ascending

# Display a fe rows for the count per category
count_business_categories.head()
```

Out[40]:

	business
category	
Banking & Finance	37
Distillers, Vintners, & Breweries	23
Aviation & Transport	19
Postal Service	16
Manufacturing & Production	15

In [46]:
```
# Create a DataFrame which lists the cumulative years businesses from each category have
count_business_years = business_categories.groupby("category").agg({"year_founded":"sum"

#Give it a new column
count_business_years.columns = ['total_years_in_business']
count_business_years = count_business_years.sort_values("total_years_in_business", ascen

# Display a few rows of the DataFrame
count_business_years.head()
```

Out[46]:

	total_years_in_business
category	
Banking & Finance	70302
Distillers, Vintners, & Breweries	39144
Aviation & Transport	36598
Postal Service	29113
Manufacturing & Production	21854

7. Restaurant Representation

No matter how we measure it, looks like Banking and Finance is an excellent industry to be in if longevity is our goal! Let's zoom in on another industry: cafés, restaurants, and bars. Which restaurants in our dataset have been around since before the year 1800?

In [47]:
```
# Filter using .query() for CAT4 businesses founded before 1800; sort results
old_restaurants = business_categories.query('category_code == "CAT4" and year_founded <
old_restaurants = old_restaurants.sort_values("year_founded", ascending=True)

# Display a few rows of the old restaurants
old_restaurants.head()
```

Out[47]:

	category_code	category	business	year_founded	country_code
65	CAT4	Cafés, Restaurants & Bars	St. Peter Stifts Kulinarium	803	AUT
66	CAT4	Cafés, Restaurants & Bars	Sean's Bar	900	IRL
62	CAT4	Cafés, Restaurants & Bars	Ma Yu Ching's Bucket Chicken House	1153	CHN

In [49]:
```
# Sort for the necessary columns
old_restaurants = old_restaurants[["business", "category", "category_code", "year_founde

# Display a few rows of the sorted old restaurants
old_restaurants.head()
```

Out[49]:

	business	category	category_code	year_founded
65	St. Peter Stifts Kulinarium	Cafés, Restaurants & Bars	CAT4	803
66	Sean's Bar	Cafés, Restaurants & Bars	CAT4	900
62	Ma Yu Ching's Bucket Chicken House	Cafés, Restaurants & Bars	CAT4	1153

8. Categories and Continents

St. Peter Stifts Kulinarium is old enough that the restaurant is believed to have served Mozart - and it would have been over 900 years old even when he was a patron! Let's finish by looking at the oldest business in each category of commerce for each continent.

In [55]:
```
# Merge business, categories and countries all together
businesses_categories_countries = business_categories.merge(countries, on="country_code"

#Display a few rows of the DataFrame
businesses_categories_countries.head()
```

Out[55]:

	category_code	category	business	year_founded	country_code	country	continent
0	CAT6	Construction	Kongō Gumi	578	JPN	Japan	Asia
1	CAT4	Cafés, Restaurants & Bars	St. Peter Stifts Kulinarium	803	AUT	Austria	Europe
2	CAT9	Distillers, Vintners, & Breweries	Staffelter Hof Winery	862	DEU	Germany	Europe
3	CAT12	Manufacturing & Production	Monnaie de Paris	864	FRA	France	Europe
4	CAT12	Manufacturing & Production	The Royal Mint	886	GBR	United Kingdom	Europe

In [56]:
```
# Sort businesses_categories_countries from oldest to most recent
businesses_categories_coutries = businesses_categories_countries.sort_values("year_found
```

```
#Display a few rows of the sorted DataFrame
businesses_categories_countries.head()
```

Out[56]:

	category_code	category	business	year_founded	country_code	country	continent
0	CAT6	Construction	Kongō Gumi	578	JPN	Japan	Asia
1	CAT4	Cafés, Restaurants & Bars	St. Peter Stifts Kulinarium	803	AUT	Austria	Europe
2	CAT9	Distillers, Vintners, & Breweries	Staffelter Hof Winery	862	DEU	Germany	Europe
3	CAT12	Manufacturing & Production	Monnaie de Paris	864	FRA	France	Europe
4	CAT12	Manufacturing & Production	The Royal Mint	886	GBR	United Kingdom	Europe

In [71]:

```
# Create the oldest by continent and category DataFrame
oldest_by_continent_categories = businesses_categories_countries.groupby(['continent', '
#oldest_by_continent_categories = oldest_by_continent_categories.sort_values("year_found

#Display the oldest_by_continent_categories
oldest_by_continent_categories.head()
```

Out[71]:

		year_founded
continent	**category**	
Africa	**Agriculture**	1947
	Aviation & Transport	1854
	Banking & Finance	1892
	Distillers, Vintners, & Breweries	1933
	Energy	1968

In [ ]: