1. The oldest businesses in the world

This is Staffelter Hof Winery, Germany's oldest business, which was established in 862 under the Carolingian dynasty. 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? Image credit: Martin Kraft

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:

`businesses` and `new_businesses`

column	type	meaning
`business`	varchar	Name of the business.
`year_founded`	int	Year the business was founded.
`category_code`	varchar	Code for the category of the business.
`country_code`	char	ISO 3166-1 3-letter country code.

`countries`

column	type	meaning
`country_code`	varchar	ISO 3166-1 3-letter country code.
`country`	varchar	Name of the country.
`continent`	varchar	Name of the continent that the country exists in.

`categories`

column	type	meaning
`category_code`	varchar	Code for the category of the business.
`category`	varchar	Description of the business category.

Now let's learn about some of the world's oldest businesses still in operation!

In [96]:

```
# Import the pandas library under its usual alias
import pandas as pd

# Load the business.csv file as a DataFrame called businesses
businesses = pd.read_csv('datasets/businesses.csv')

# Sort businesses from oldest businesses to youngest
sorted_businesses = businesses.sort_values(['year_founded'],ascending=True)

# Display the first few lines of sorted_businesses
sorted_businesses.head()
```

Out[96]:

	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

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?

Enter `countries.csv`, which is also located in the `datasets` folder. Having useful information in different files is a common problem: for data storage, it's better to keep different types of data separate, but for analysis, we want all the data in one place. To solve this, we'll have to join the two tables together.

`countries`

column	type	meaning
`country_code`	varchar	ISO 3166-1 3-letter country code.
`country`	varchar	Name of the country.
`continent`	varchar	Name of the continent that the country exists in.

Since `countries.csv` contains a `continent` column, merging the datasets will also allow us to look at the oldest business on each continent!

In [98]:

```
# Load countries.csv to a DataFrame
countries = pd.read_csv('datasets/countries.csv')

# Merge sorted_businesses with countries
businesses_countries = sorted_businesses.merge(countries,on='country_code')

# Filter businesses_countries to include countries in North America only
north_america = businesses_countries[businesses_countries['continent'] == 'North America']
north_america.head()
```

Out[98]:

	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 business on each continent

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

In [100…

```
# Create continent, which lists only the continent and oldest year_founded
continent = businesses_countries.groupby('continent').agg({'year_founded':'min'})

# Merge continent with businesses_countries
merged_continent = continent.merge(businesses_countries,on=['continent','year_founded'])

# Subset continent so that only the four columns of interest are included
subset_merged_continent = merged_continent[['continent','country','business','year_founded']]
subset_merged_continent
```

Out[100…

	continent	country	business	year_founded
0	Africa	Mauritius	Mauritius Post	1772
1	Asia	Japan	Kongō Gumi	578
2	Europe	Austria	St. Peter Stifts Kulinarium	803
3	North America	Mexico	La Casa de Moneda de México	1534
4	Oceania	Australia	Australia Post	1809
5	South America	Peru	Casa Nacional de Moneda	1565

4. Unknown oldest 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 [102…

```
# Use .merge() to create a DataFrame, all_countries
all_countries = businesses.merge(countries, on="country_code", how="right",  indicator=True)

# Filter to include only countries without oldest businesses
missing_countries = all_countries[all_countries["_merge"] != "both"]

# Create a series of the country names with missing oldest business data
missing_countries_series = missing_countries["country"]

# Display the series
missing_countries_series
```

Out[102…

```
163                              Angola
164                 Antigua and Barbuda
165                             Bahamas
166                  Dominican Republic
167                             Ecuador
168                                Fiji
169     Micronesia, Federated States of
170                               Ghana
171                              Gambia
172                             Grenada
173           Iran, Islamic Republic of
174                          Kyrgyzstan
175                            Kiribati
176               Saint Kitts and Nevis
177                              Monaco
178                Moldova, Republic of
179                            Maldives
180                    Marshall Islands
181                               Nauru
182                               Palau
183                    Papua New Guinea
184                            Paraguay
185                Palestine, State of
186                     Solomon Islands
187                            Suriname
188                          Tajikistan
189                        Turkmenistan
190                         Timor-Leste
191                               Tonga
192                              Tuvalu
193    Saint Vincent and the Grenadines
194                               Samoa
Name: country, dtype: object
```

5. Adding new oldest business 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 `"datasets/new_businesses.csv"`. It has the exact same structure as our `businesses` dataset.

`new_businesses`

column	type	meaning
`business`	varchar	Name of the business.
`year_founded`	int	Year the business was founded.
`category_code`	varchar	Code for the category of the business.
`country_code`	char	ISO 3166-1 3-letter country code.

All we have to do is combine the two so that we've got one more complete list of businesses!

In [104…

```
# Import new_businesses.csv
new_businesses = pd.read_csv('datasets/new_businesses.csv')

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

# Merge and filter to find countries with missing business data
new_all_countries = all_businesses.merge(countries, on='country_code', how='outer', indicator=True)
new_missing_countries = new_all_countries[new_all_countries["_merge"] != "both"]

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

Out[104…

continent	count_missing
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?

	business	year_founded	category_code	country_code
64	Kongō Gumi	578	CAT6	JPN

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 `"datasets/categories.csv"`:

`categories`

column	type	meaning
`category_code`	varchar	Code for the category of the business.
`category`	varchar	Description of the business category.

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

In [106…

```
# Import categories.csv and merge to businesses
categories = pd.read_csv("datasets/categories.csv")
businesses_categories = businesses.merge(categories, on="category_code")

# Create a DataFrame which lists the number of oldest businesses in each category
count_business_cats = businesses_categories.groupby("category").agg({"business":"count"})

# Create a DataFrame which lists the cumulative years that businesses from each category have been operating
years_business_cats = businesses_categories.groupby("category").agg({"year_founded":"sum"})

# Rename columns and display the first five rows of both DataFrames
count_business_cats.columns = ["count"]
years_business_cats.columns = ["total_years_in_business"]
display(count_business_cats.head(), years_business_cats.head())
```

category	count
Agriculture	6
Aviation & Transport	19
Banking & Finance	37
Cafés, Restaurants & Bars	6
Conglomerate	3

category	total_years_in_business
Agriculture	10669
Aviation & Transport	36598
Banking & Finance	70302
Cafés, Restaurants & Bars	8532
Conglomerate	5671

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 [108…

```
# Filter using .query() for CAT4 businesses founded before 1800; sort results
old_restaurants = businesses_categories.query('year_founded < 1800 and category_code == "CAT4"')

# Sort the DataFrame
old_restaurants = old_restaurants.sort_values("year_founded")
old_restaurants
```

Out[108…

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

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 [110…

```
# Merge all businesses, countries, and categories together
businesses_categories_countries = businesses_categories.merge(countries, on="country_code")

# Sort businesses_categories_countries from oldest to most recent
businesses_categories_countries = businesses_categories_countries.sort_values("year_founded")

# Create the oldest by continent and category DataFrame
oldest_by_continent_category = businesses_categories_countries.groupby(["contine
oldest_by_continent_category.head()
```

Out[110…

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