**Enduring Enterprises: Analyzing the World’s Oldest Businesses**

Anesu David Pasipanodya

All the analysis are based on data in the project and the opinions of the author as of the 22nd of July 2024

<https://github.com/AnesuorDavid/AnesuorDavid.github.io>

Abstract

This project investigates the remarkable longevity of businesses across the globe by analyzing the oldest currently operating company in nearly every country. Using a dataset compiled by BusinessFinancing.co.uk, which includes information on founding year, business type, and location, the analysis aims to identify patterns and common traits among the world's longest-standing enterprises. By joining and manipulating the provided datasets, we explore continental trends, industry distributions, and historical resilience, offering insights into what enables businesses to endure for centuries. This study also considers the impact of geography, economic environment, and business category on long-term survival.

Contents

[Contents 3](#_Toc204098853)

[Introduction 4](#_Toc204098854)

[Defining the Project 4](#_Toc204098855)

[Pre-Analysing the Dataset 5](#_Toc204098856)

[Data Uploading 6](#_Toc204098857)

[Data Normalisation and Cleaning 6](#_Toc204098858)

[Data aggregation 8](#_Toc204098859)

Introduction

Longevity in the business world is a rare and impressive achievement. Few companies withstand the test of time, weathering political upheaval, economic crises, technological disruptions, and cultural change. Staffelter Hof, a German winery established in 862, is a prime example of such resilience—surviving the rise and fall of empires, global wars, and shifting marketplaces while continuing to operate to this day.

This project explores the oldest existing businesses on each continent and in each country, drawing on data curated by BusinessFinancing.co.uk. The dataset includes foundational information such as business names, founding years, country and continent classifications, and industry categories. The data is on a single CSV file the data, which show organizational clarity, meaningful analysis requires combining and transforming these sources into a cohesive structure.

Through structured querying and data visualization, the goal of this project is to uncover what types of businesses tend to last the longest, which regions foster such longevity, and how different industries fare over centuries. This analysis not only highlights the incredible endurance of certain enterprises but also provides valuable lessons for modern businesses seeking long-term sustainability.

Defining the Project

“In order to get insights from the data we have to ask the right questions.”

**Analysis of the Oldest Registered Businesses by Country**

This project investigates the characteristics and longevity of some of the world’s oldest continuously operating businesses. By examining data points such as founding year, industry category, and operational status, the project aims to identify trends and patterns that contribute to exceptional business durability.

**The central goal** is to uncover which types of businesses tend to endure across centuries, what historical patterns of business formation emerge over time, and how age correlates with sector or type of service.

## Objectives:

* Understand what kinds of businesses have stood the test of time.
* Explore how the rate of business creation has evolved across historical periods.
* Identify the common traits (e.g., industry type, founding era) shared by century-old businesses.
* Determine the distribution of founding years and business lifespans in the dataset.
* Avoid region- or country-specific analysis and focus solely on time and business characteristics.

In order to analyze the historical and geographical distribution of the oldest businesses worldwide, we will ask the following questions.

1. Which business categories are best suited to last over the course of centuries?
2. What is the oldest business in the dataset overall?
3. How has the rate of new business founding changed over time?
4. What’s the distribution of founding years across the dataset?
5. Which industries or sectors contain the highest number of century-old businesses?
6. What is the average founding year per business category?
7. Are there any clusters of businesses founded during significant historical periods (e.g., 1850–1900, post-WWII)?
8. How many businesses still operating today were founded before 1800?
9. Are there any categories that show consistent business formation across multiple centuries?
10. Which business category has the most businesses founded in the 20th century?

Pre-Analysing the Dataset

The dataset consists of a .csv file detailing the oldest still-operating businesses across the globe, compiled by **BusinessFinancing.co.uk**. It includes business names, founding years, country and category codes, along with separate lookup tables for countries and business categories. To conduct the analysis, these files were imported into a relational database using MySQL, where joins were performed to unify the data across business, country, and category tables. This enabled a more holistic view of each business’s location, continent, and industry type. Data cleaning processes ensured consistency in formats and removed any anomalies or duplicate entries. Aggregation and filtering techniques were then applied to identify key trends, such as the oldest business on each continent, regional longevity patterns, and dominant industries among long-standing companies. Visual exploration using Tableau helped bring the analysis to life, revealing insightful geographical and temporal patterns that show how businesses have endured over centuries.

Data Uploading

I imported the data by creating a schema in MySQL. Then I added the added to the table data import wizard.

**QUERY**

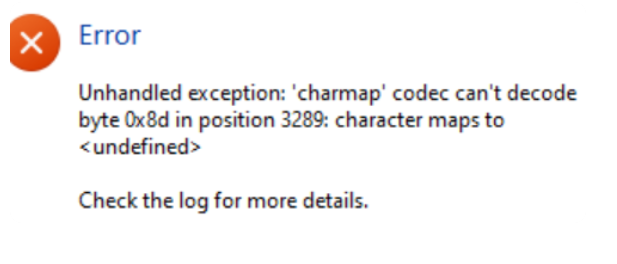
|  |
| --- |
| -- **DATA IMPORT**  SELECT \*  FROM businesses  LIMIT 5; |

**ANSWER**

|  |  |  |  |
| --- | --- | --- | --- |
| # ï»¿index | business | year\_founded | category\_code |
| 0 | Hamoud Boualem | 1878 | CAT11 |
| 1 | Communaute Electrique du Benin | 1968 | CAT10 |
| 2 | Botswana Meat Commission | 1965 | CAT1 |
| 3 | Air Burkina | 1967 | CAT2 |
| 4 | Brarudi | 1955 | CAT9 |

Data Normalisation and Cleaning

Upon opening the data in MySQL the following error is seen:

This because the accurate names are used for the companies, and this being international companies from different countries, many special letters e.g. accents and Nordic Ø had to be changed in order for MySQL to be able to read the data. These had to be edited in MySQL.

* Then I added it to MySQL and made a staging table called enduring\_businesses where I checked if it was present. I also checked for repeated values.

**QUERY**

|  |
| --- |
| **-- STAGING table**  CREATE TABLE enduring\_businesses  LIKE businesses;  INSERT INTO enduring\_businesses  SELECT \*  FROM businesses;  SELECT \*  FROM enduring\_businesses  LIMIT **5**; |

**ANSWER**

|  |  |  |  |
| --- | --- | --- | --- |
| # ï»¿index | business | year\_founded | category\_code |
| 0 | Hamoud Boualem | 1878 | CAT11 |
| 1 | Communaute Electrique du Benin | 1968 | CAT10 |
| 2 | Botswana Meat Commission | 1965 | CAT1 |
| 3 | Air Burkina | 1967 | CAT2 |
| 4 | Brarudi | 1955 | CAT9 |

* Then I checked for null values and blank spaces.

**QUERY**

|  |
| --- |
| **-- CHECKING FOR BLANK SPACES AND NULL VALUES**  SELECT `ï»¿index`, business, year\_founded, category\_code, country\_code  FROM businesses  WHERE `ï»¿index` IS NULL OR `ï»¿index` = ''  AND business IS NULL OR business = ''  AND year\_founded IS NULL OR year\_founded = ''  AND category\_code IS NULL OR category\_code = ''  AND country\_code IS NULL OR country\_code = ''  GROUP BY `ï»¿index`, business, year\_founded, category\_code, country\_code; |

* There were 0 results

ANSWER

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ï»¿index | business | year\_founded | category\_code | country\_code |
|  |  |  |  |  |

* I also checked for repeated values.

**QUERY**

|  |
| --- |
| **-- CHECKING FOR REPEATED VALUES**  SELECT `ï»¿index`, business, year\_founded, category\_code, country\_code, COUNT(\*) AS occurrences  FROM businesses  GROUP BY `ï»¿index`, business, year\_founded, category\_code, country\_code  HAVING COUNT(\*) > 1; |

**ANSWER**

* No repeated values

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ï»¿index | business | year\_founded | category\_code | country\_code |
|  |  |  |  |  |

Tableau and Excel.

To further analyse data I also uploaded data sets via copy and paste from query results into excel where I added them to Tableau for further analysis

Data Analysis through Aggregation

1. *Which business categories are best suited to last over the course of centuries?*

To calculate which categories are best suited to last over the course of centuries, we have to separate the business categories by the ones that appear the most.

**QUERY**

|  |
| --- |
| SELECT category\_code, COUNT(\*) AS freq  FROM enduring\_businesses  GROUP BY category\_code  ORDER BY freq DESC  LIMIT 5; |

**ANSWER**

|  |  |
| --- | --- |
| category\_code | freq |
| CAT3 | 148 |
| CAT9 | 88 |
| CAT2 | 76 |
| CAT16 | 64 |
| CAT12 | 60 |

From here we can see which have the most over centuries. We can take different approaches – we can see which category has the most businesses still in operation, or we can see which has the oldest business in general. As a result of us only seeking which industry has longevity we will only look at year\_founded.

QUERY+ ANSWERS

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SELECT category\_code, MIN(year\_founded) AS earliest\_year  FROM enduring\_businesses  GROUP BY category\_code  ORDER BY earliest\_year ASC;   |  |  | | --- | --- | | category\_code | earliest\_year | | CAT9 | 862 | | CAT12 | 864 | | CAT16 | 1520 | | CAT3 | 1565 | | CAT2 | 1854 | |

Based on the data, CAT9 is the oldest category with the most frequency. The data reveals that these are distilleries, breweries and liquor producers. The oldest being StaffelterÂ Hof Winery from Germany (DEU), its earliest year being in 862CE.

1. *What is the oldest business in the dataset overall?*

The oldest business is easy to see. We just query the earliest year of a business and its business name.

**QUERY**

|  |
| --- |
| -- 2. What is the oldest business in the dataset overall?  SELECT business, year\_founded  FROM enduring\_businesses  GROUP BY business, year\_founded  ORDER BY year\_founded  limit 1; |

**ANSWER**

|  |  |
| --- | --- |
| **business** | **year\_founded** |
| Kongo Gumi | 578 |

The Kongo Gumi is a n almost 1500-year-old construction company and it is the oldest in the given data

1. *How has the rate of new business founding changed over time?*

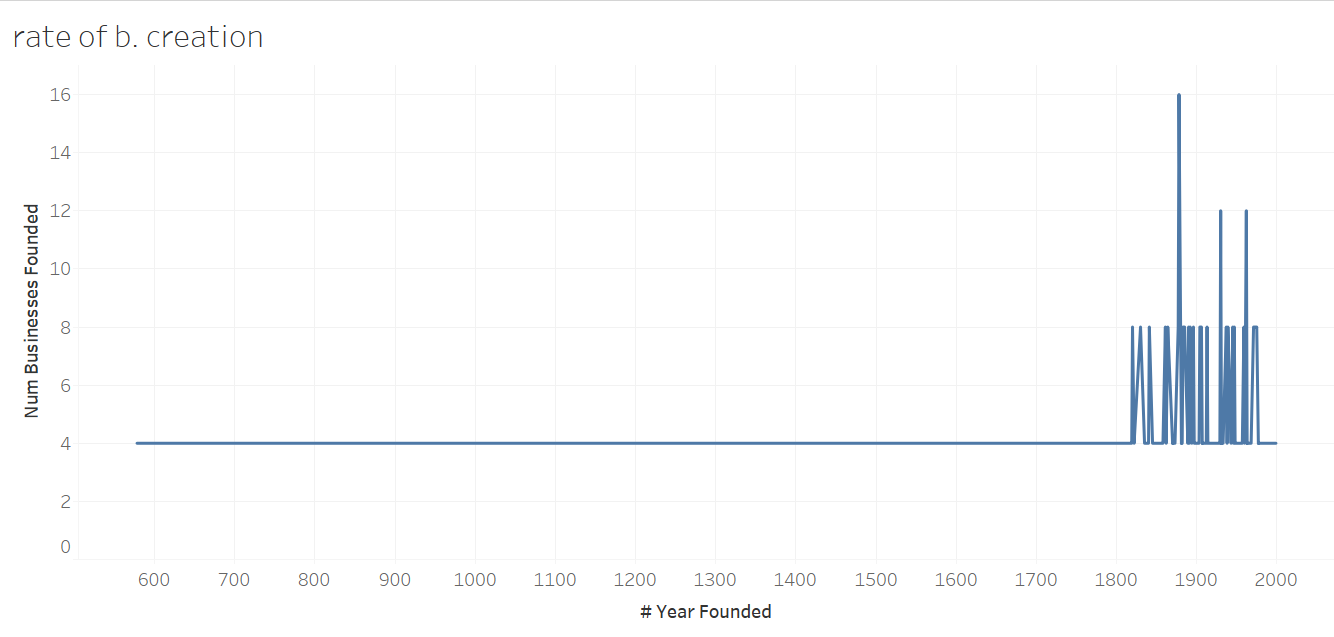
To analyse rate of business\_founding over time we use the year\_founded field from the enduring\_businesses dataset to count how many businesses were founded each year. Then we group the data by year to calculate the number of new businesses founded annually. We export those tables to excel to visualise the results to observe trends across decades or centuries.

**QUERY**

|  |
| --- |
| SELECT year\_founded, COUNT(\*) AS num\_businesses\_founded  FROM enduring\_businesses  GROUP BY year\_founded  ORDER BY year\_founded ASC; |

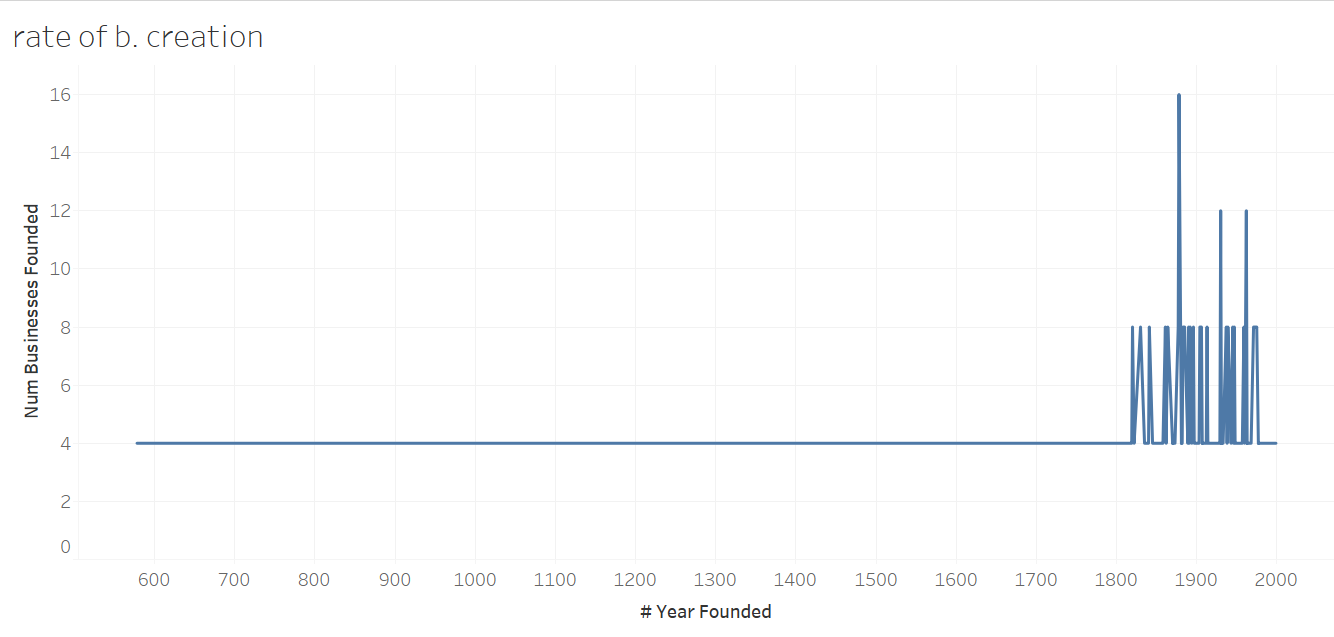
ANSWER (rateofnewbusiness.xsl)

|  |  |
| --- | --- |
| # year\_founded | num\_businesses\_founded |
| 578 | 4 |
| 803 | 4 |
| 862 | 4 |
| 864 | 4 |
| 886 | 4 |
| 900 | 4 |
| 1040 | 4 |
| 1074 | 4 |
| 1135 | 4 |
| 1153 | 4 |
| 1218 | 4 |
| 1230 | 4 |
| 1248 | 4 |
| 1250 | 4 |
| 1328 | 4 |

On the x-axis there is the Year Founded, which range from the mid 500’s to the early 2000’s and on the y axis has number of businesses founded in a year. Upon analysis of the above graph there is a constant founding rate of 4 businesses a year from the 6th Century to the early 19th century. From the year of 578 the rate does not drop to lower than this rate of 4 businesses per year. In the 19th century it fluctuated greatly, possibly between the great industrial advances of that century amid all the wars of that century. It had peaks in 1820, 1830, 1841, 1861 and 1864 where 8 companies were made that still existed till the creation of the dataset. As it approached the beginning of the 20th century it spiked heavily especially in 1878, where 16 companies being made per. It dropped to the standard rate of 4 companies a year in 1882. It had the peaks of 8 companies a year from 1881 – 1885, then 1890, 1892 and 1896. It dropped to standard 4 a year till 1902, where it spiked back to that rate from 1904 – 1906. From 1907 to 1912 it dropped back to 4 businesses a year then rose back to 8b/yr in 1913. From 1914 to 1929 the normal rate of 4 b/yr till the year of the Great Depression. It then peaked to 12b/yr where it dropped back the next year. It maintained this till 1933 where it peaked to 8 b/yr in 1937, 1939, 1945, 1947, and 1959 respectively. The rate dropped back to 4 b/yr in 1960, then shot up the very next year in 1962 to 12b/yr. in 1963 the rate dropped back to the standard 4b/yr rate of starting business. From 1971 to 1975 8 businesses were founded per year, then from 1977 onwards the rate returned to 4b/yr.

1. *What’s the distribution of founding years across the dataset?*

To understand the distribution of founding years, we examine how business founding dates are spread across time in the dataset. Upon observing the graph below the distribution is strongly concentrated in the last 200 years, with peaks in major war, and historical periods. Very few businesses date back before the 1700s, indicating the rarity of extremely old, enduring companies. The distribution is right-skewed, meaning most businesses are relatively modern, with a long tail of very old businesses. This means that according to this dataset, most long-lasting businesses were founded in more recent centuries, reflecting the growth of commerce and industrialization. The small number of very old businesses suggests rarity and resilience, highlighting how few businesses survive for centuries. The pattern may reflect data availability bias (e.g., more records for recent years), or real-world factors like global economic expansion in the modern era.



1. *Which industries or sectors contain the highest number of century-old businesses?*

As seen in the first question, we can easily aggregate the data through previously used queries that highlight Categories with the greatest number of businesses across the centuries. This is shown with the queries below:

**QUERY**

|  |
| --- |
| SELECT category\_code, COUNT(\*) AS freq  FROM enduring\_businesses  GROUP BY category\_code  ORDER BY freq DESC  LIMIT 5; |

**ANSWER**

|  |  |
| --- | --- |
| category\_code | freq |
| CAT3 | 148 |
| CAT9 | 88 |
| CAT2 | 76 |
| CAT16 | 64 |
| CAT12 | 60 |

Upon glancing at this result Category 3 has the highest frequency of businesses spanning across the centuries. To see what kind of companies are considered ‘CAT3’, we only have to query all CAT3 companies.

**QUERIES:**

|  |
| --- |
| SELECT \*  FROM enduring\_businesses  WHERE category\_code = 'CAT3'  LIMIT 5; |

**ANSWER**:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | ï»¿index | business | year\_founded | category\_code | country\_code | | 7 | Banque Internationale pour la Centrafrique | 1946 | CAT3 | CAF | | 9 | Central Bank of the Comoros | 1981 | CAT3 | COM | | 11 | Development Bank of the Central African States | 1975 | CAT3 | COG | | 17 | National Bank of Ethiopia | 1906 | CAT3 | ETH | |

As can easily be seen **banks** have the greatest number of century long businesses.

1. *What is the average founding year per business category?*

This is done by finding the average year of each category code.

**QUERY:**

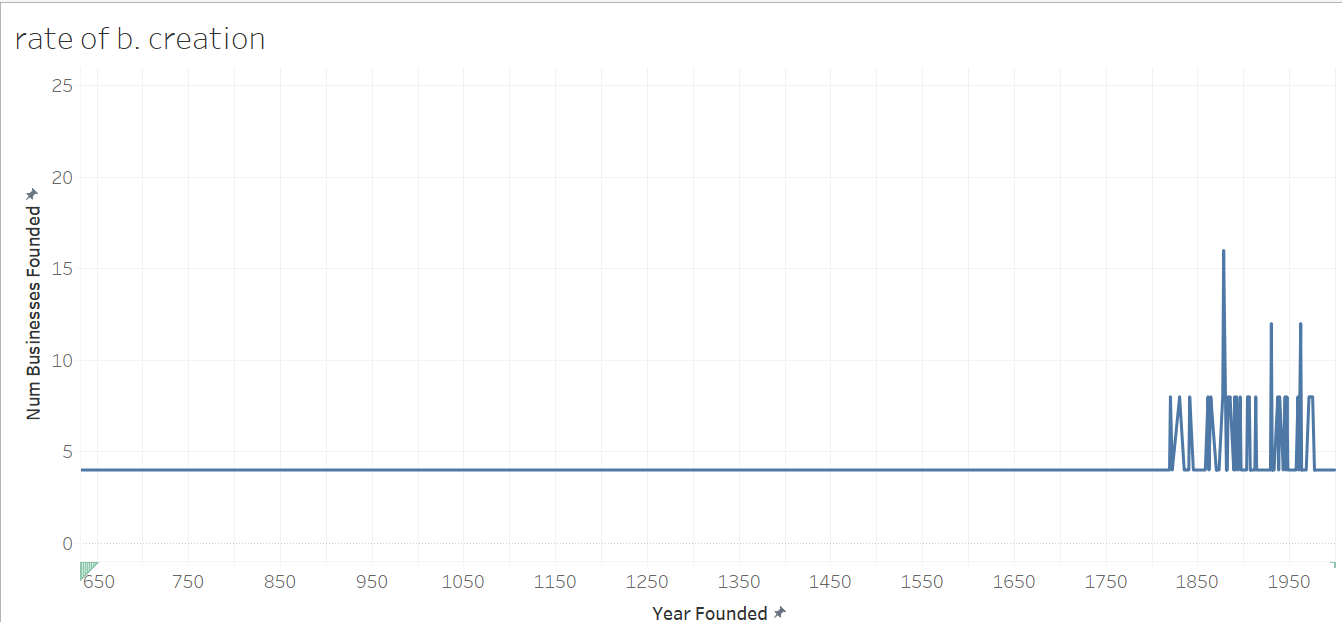
|  |
| --- |
| SELECT category\_code, ROUND(AVG(year\_founded),0) AS Avg\_founding\_date  FROM enduring\_businesses  GROUP BY category\_code; |

**ANSWER**:

|  |  |
| --- | --- |
| # category\_code | Avg\_founding\_date |
| CAT13 | 1952 |
| CAT10 | 1948 |
| CAT2 | 1926 |
| CAT3 | 1900 |
| CAT18 | 1899 |
| CAT5 | 1890 |
| CAT17 | 1848 |
| CAT8 | 1832 |
| CAT11 | 1826 |
| CAT16 | 1820 |
| CAT1 | 1778 |
| CAT7 | 1748 |
| CAT15 | 1708 |
| CAT9 | 1695 |
| CAT19 | 1513 |
| CAT12 | 1457 |
| CAT4 | 1422 |
| CAT14 | 1422 |
| CAT6 | 1212 |

* The data shows that **CAT6** (**Construction of religious buildings**) are the oldest companies on average followed by **CAT14**, **CAT4**, **CAT12** and **CAT19**. These are **ancient pharmacies**, **restaurants**, **manufacturers** and **hotels/hospitality** respectively.
* In contrast the youngest companies on average are **CAT13, CAT10, CAT2, CAT3** and **CAT18**. These are **Broadcasting** **corporations**, **Energy companies, Airlines, National banks and telecom companies** respectively with Broadcasting corporations being the youngest industry.
* This data shows that health, hospitality, restaurants, and manufacturers are long lasting industries that regardless of age humans will always need.

1. *Are there any clusters of businesses founded during significant historical periods (e.g., 1850–1900, post-WWII)?*

Yes. To prove this we will only use the graph on Tableau.

The dataset reveals distinct clusters of business founding during several historically significant periods, based on the trends observed in the founding years.

**1850–1900 (Industrial Revolution Peak):**

There is a noticeable increase in business founding in this period. Peaks occur in 1861, 1864, 1878, and 1890s, with up to 16 businesses founded in a single year (1878). This clustering aligns with the second wave of the Industrial Revolution, marked by advances in steel, railways, and electricity — enabling large-scale commerce and manufacturing.

**Post-WWI & Interwar Period (1918–1939):**

After a dip in WWI, founding rates begin to rise again, particularly around 1929–1933, despite the Great Depression. This may suggest businesses that weathered or responded to economic shocks with innovation and adaptability.

**Post-WWII (1945–1960):**

There’s another spike in founding rates during 1945, 1947, and 1959, reflecting a post-war economic boom. This period saw global reconstruction, trade liberalization, and the growth of consumer goods markets — all of which created fertile ground for new business formation.

**Early 1960s & 1970s:**

Notable spikes in 1962 (12 businesses), and again in 1971–1975 (8 businesses/year). These clusters may relate to the globalization wave, technological advancement, and economic reforms in many developed nations.

1. *How many businesses still operating today were founded before 1800?*

We will query this by seeing the number of businesses made > 1800

**QUERY**:

|  |
| --- |
| SELECT COUNT(\*)  FROM enduring\_businesses  WHERE year\_founded > 1800; |

**ANSWER**

|  |
| --- |
| 472 OUT OF 652 |

Here we can see through basic calculation that **72.3**% of all the companies in the dataset where made after the year of 1800.

1. *Are there any categories that show consistent business formation across multiple centuries?*

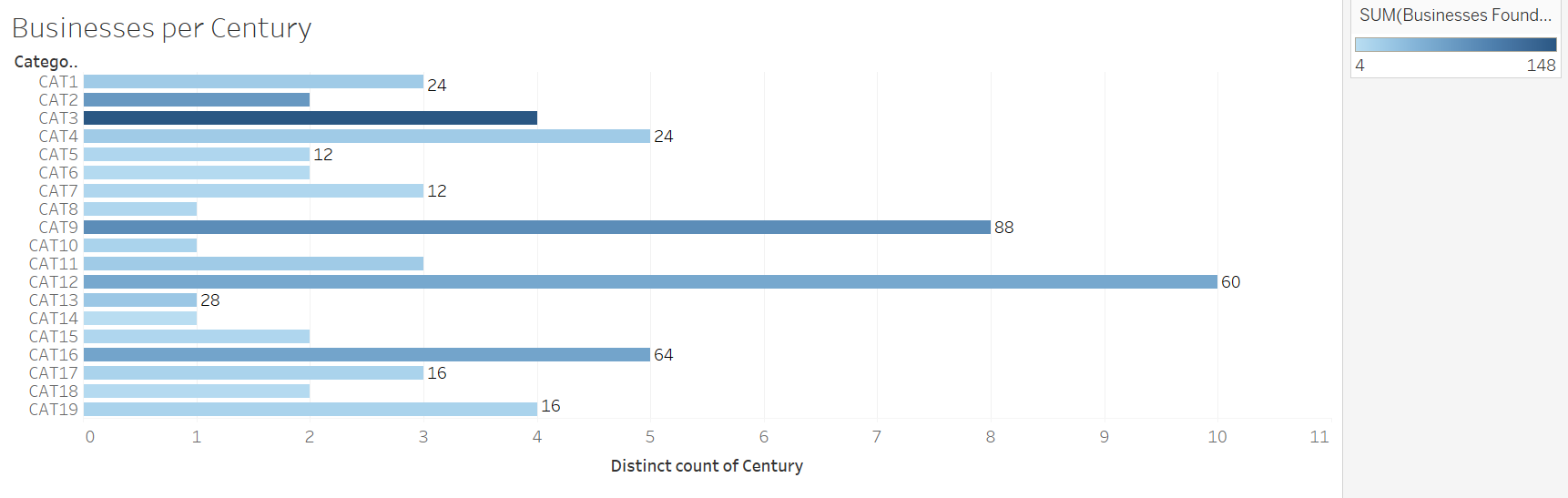
To analyse this we would have to categorise data by century. We will then have to put these in Excel and Tableau for further analysis

**QUERY**:

|  |
| --- |
| SELECT  category\_code,  FLOOR(year\_founded / 100) + 1 AS century,  COUNT(\*) AS businesses\_founded  FROM enduring\_businesses  GROUP BY category\_code, century  ORDER BY category\_code, century  LIMIT 5; |

**ANSWER:**

|  |  |  |
| --- | --- | --- |
| category\_code | century | businesses\_founded |
| CAT1 | 13 | 4 |
| CAT1 | 17 | 4 |
| CAT1 | 20 | 16 |
| CAT10 | 20 | 16 |
| CAT11 | 17 | 4 |

I exported the full tables from the SQL query to excel. I then inserted the excel file into Tableau and used a heatmap text table in order to analyse it effectively.

The way this works is that the x-axis shows the number of centuries the industry categories span. The number at the end of the bars show the number of businesses started in those centuries. The Category Number is the industry spoken about throughout the project.

* Analysis shows Category 12 has had 60 businesses created in over 10 centuries 9th, 11th, 12th, 13th, 14th, 16th, 17th, 18th, 19th and 20th centuries. Curiously the industry is manufacturing and although one would expect to peak in later years, it has been consistent over the centuries.

1. *Which business category has the most businesses founded in the 20th century?*

We will analyse data from 1901 – 2000 with an SQL query. It will help us analyse modern growth trends

**QUERY**:

|  |
| --- |
| SELECT  category\_code,  COUNT (\*) AS businesses\_founded\_20th  FROM enduring\_businesses  WHERE year\_founded BETWEEN 1901 AND 2000  GROUP BY category\_code  ORDER BY businesses\_founded\_20th DESC; |

**ANSWER:**

|  |  |
| --- | --- |
| # category\_code | businesses\_founded\_20th |
| CAT3 | 92 |
| CAT2 | 52 |
| CAT13 | 28 |
| CAT10 | 16 |
| CAT1 | 16 |
| CAT9 | 16 |
| CAT16 | 12 |
| CAT15 | 8 |
| CAT12 | 8 |
| CAT17 | 4 |
| CAT4 | 4 |
| CAT5 | 4 |
| CAT18 | 4 |
| CAT11 | 4 |

The categories with the highest number of businesses founded in the 20th century are **Financial Institutions (CAT3)**, **Airlines (CAT2)**, and **Broadcasting Corporations (CAT13)**. This suggests these sectors experienced significant growth during the modern era — likely driven by industrialisation, the rise of commercial air travel, and the global expansion of mass media. These industries grew rapidly in the 20th century due to major global changes:

**Financial Institutions** expanded to support industrialisation, global trade, and consumer banking.

**Airlines** emerged with advances in aviation, connecting countries and boosting travel and commerce.

**Broadcasting Corporations** rose with the invention of radio and television, creating demand for media and entertainment.

All three reflect how technology, globalization, and modern infrastructure shaped new business opportunities.

Acknowledgements

DataCamps – for the dataset, and project ideas

BusinessFinancing.co.uk