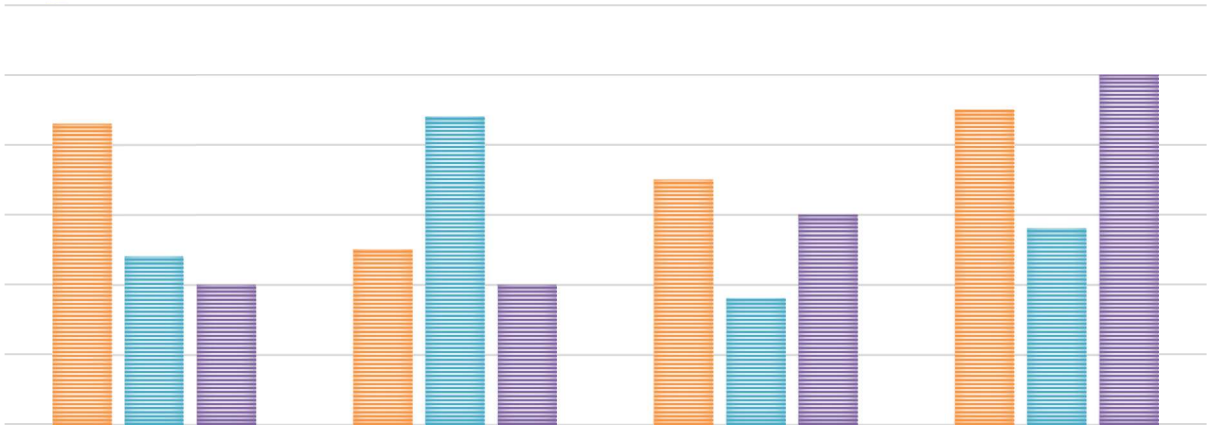
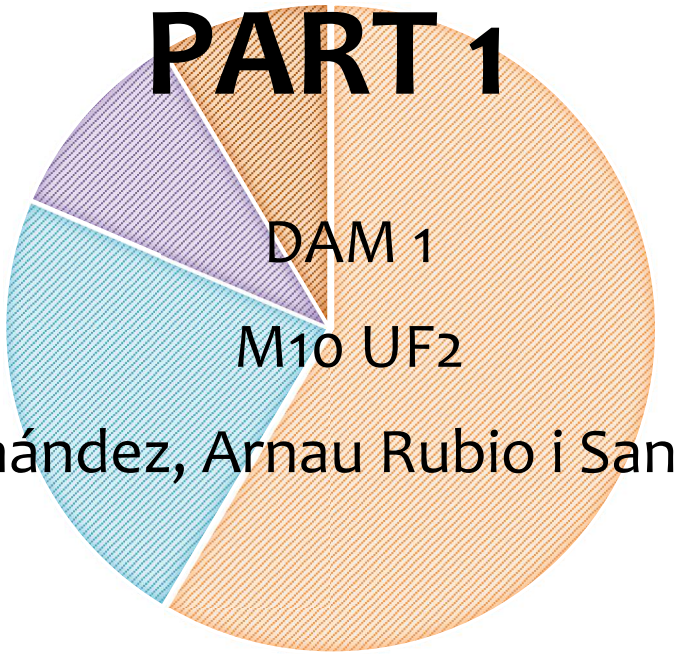


PROJECTE FINAL

PART 1



DAM 1

M10 UF2

Ivan Fernández, Arnau Rubio i Sandra Rubio

Projecte Final
Sandra Rubio Poveda
Arnau Rubio Poveda
Ivan Fernández Montalvo

ÍNDEX

1. Introducció

A partir d'una base de dades sobre les persones més milionàries del món, realitzarem diferents estadístiques per determinar certs paràmetres i poder obtenir informació interessant. Com ara:

- La distribució de l' edat dels milionaris, per poder veure l'edat que tenen la majoria de milionaris.
- Els països amb més quantitat de milionaris residents i la distribució per país de residencia i d'origen
- La distribució del gènere dels milionaris.
- El sector d'on prové la riquesa dels milionaris i l distribució d'aquesta riquesa per industria.
- El percentatge de milionaris autodidactes en front aquells que han heretat la seva riquesa.
- Relació entre el PIB del país i la quantitat de milionaris
- Relació entre la tassa de matrícules d'educació primària i terciària i la quantitat de milionaris
- La distribució dels milionaris per país i ciutat.

2. Datasets seleccionats

La nostra base de dades consta d'una taula indexada pel rang (*rank*) on trobem els següents camps:

- Riquesa final. (*finalWorth*)
- Categoria de la Industria a la que pertany. (*category*)
- El nom complet de la persona. (*personName*)
- L'edat. (*age*)
- El país de residencia. (*country*)
- La ciutat de residencia. (*city*)
- La font o l'origen de la riquesa o la fama d'una persona o entitat. (*source*)
- La industria (igual que el segon camp). (*industries*)
- El país d'origen. (*countryOfCitizenship*)
- La organització a la que pertany l'empresa. (*organization*)
- Indica si el milionari ha estat autodidacta o ha heretat la riquesa. (*selfMade*)
- L'estatus social de la persona (família humil, adinerada, estudis…). (*status*)
- El gènere de la persona. (*gender*)
- La data de naixement. (*birthDate*)
- El primer cognom. (*lastName*)
- El nom. (*firstName*)
- El títol utilitzat per a una persona, com ara "Mr.", "Mrs.", "Dr."... (*title*)
- La data en que es va realitzar la base de dades. (*date*)
- L'estat d'Estats Units on viu el milionari (si no es d'EEUU és NULL). (*state*)
- La regió d'Estats Units on viu el milionari (si no es d'EEUU és NULL). (*residenceStateRegion*)

- L' any de naixement (*birthYear*).
- El mes de naixement (*birthMonth*).
- El dia de naixement (*birthDay*).
- L'índex de percepció de la corrupció del país (*cpi_country*).
- El canvi de l'índex de percepció de la corrupció del país (*cpi_change_country*).
- El producte interior brut del país de residencia (*gdp_country*).
- Matriculacions en educació terciària en percentatge (*gross_tertiary_education_enrollment*).
- Matriculacions en educació primària en percentatge (*gross_primary_education_enrollment_country*).
- L'esperança de vida (*life_expectancy_country*).
- Els ingressos fiscals per a un país en concret (*tax_revenue_country_country*).
- Taxa impositiva total per a un país en concret(*total_tax_rate_country*).
- La població del país (*population_country*).
- La latitud geogràfica del país (*latitude_country*).
- La longitud geogràfica del país (*longitude_country*).

3. Modificacions efectuades sobre els Datasets

Actualment els únics camps que no creiem utilitzar són els següents:

- La latitud geogràfica del país (*latitude_country*)
- La longitud geogràfica del país (*longitude_country*)

Per tant, els eliminarem de la nostra base de dades.

Amb les següents instruccions eliminem les dues columnes de les que volem prescindir:

```
columnas_a_eliminar = ['latitude_country', 'longitude_country']

df = df.drop(columnas_a_eliminar, axis=1)
```

Fem la següent instrucció perquè ens mostri la capçalera:

```
print(df.head())
```

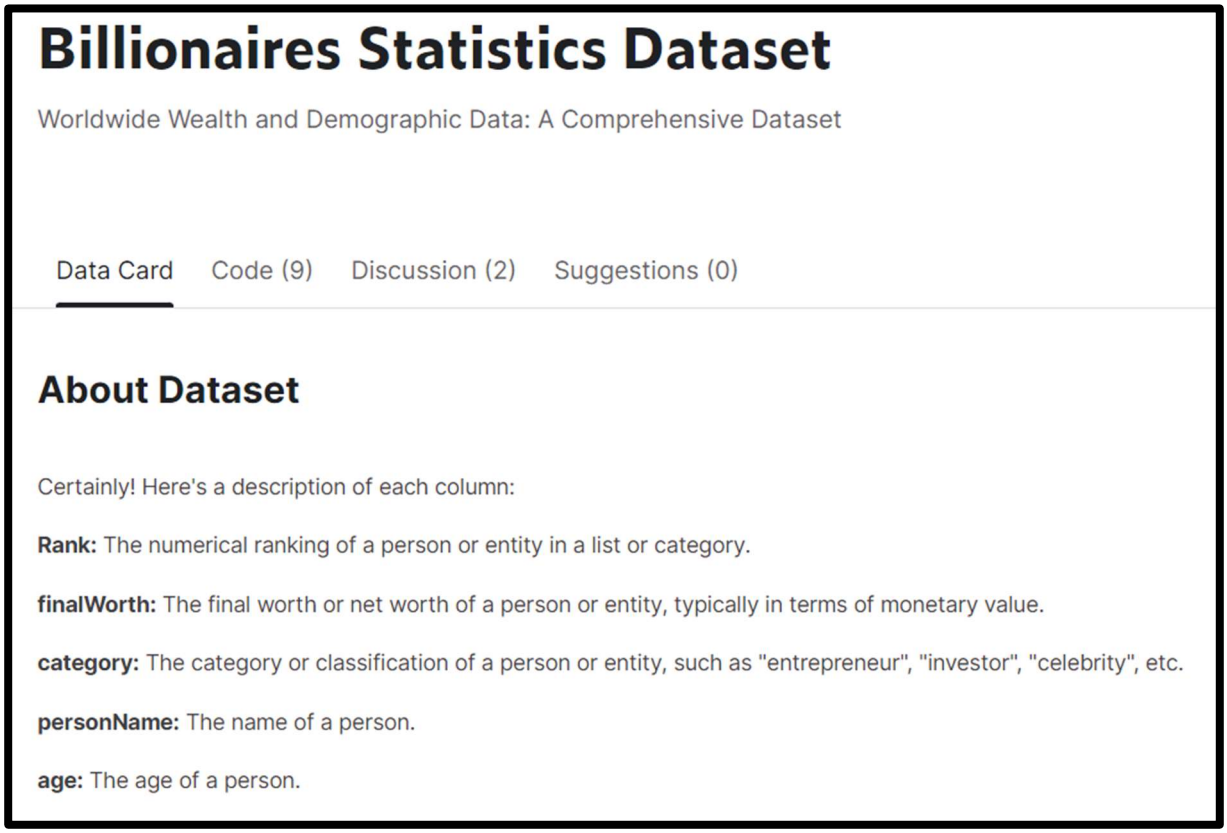
```
   rank  finalWorth               category                personName   age  \
0     1      211000       Fashion & Retail  Bernard Arnault & family  74.0
1     2      180000             Automotive                 Elon Musk  51.0
2     3      114000             Technology                Jeff Bezos  59.0
3     4      107000             Technology              Larry Ellison  78.0
4     5      106000  Finance & Investments             Warren Buffett  92.0

         country    city              source             industries  \
0         France   Paris                LVMH       Fashion & Retail
1  United States  Austin       Tesla, SpaceX             Automotive
2  United States  Medina              Amazon             Technology
3  United States   Lanai              Oracle             Technology
4  United States   Omaha  Berkshire Hathaway  Finance & Investments

  countryOfCitizenship  ... birthDay  cpi_country cpi_change_country  \
0               France  ...      5.0       110.05                1.1
1        United States  ...     28.0       117.24                7.5
2        United States  ...     12.0       117.24                7.5
3        United States  ...     17.0       117.24                7.5
4        United States  ...     30.0       117.24                7.5

           gdp_country gross_tertiary_education_enrollment  \
0   $2,715,518,274,227                                65.6
1  $21,427,700,000,000                                88.2
2  $21,427,700,000,000                                88.2
3  $21,427,700,000,000                                88.2
4  $21,427,700,000,000                                88.2

  gross_primary_education_enrollment_country life_expectancy_country  \
0                                      102.5                    82.5
1                                      101.8                    78.5
2                                      101.8                    78.5
3                                      101.8                    78.5
4                                      101.8                    78.5

  tax_revenue_country_country total_tax_rate_country population_country
0                        24.2                   60.7           67059887.0
1                         9.6                   36.6          328239523.0
2                         9.6                   36.6          328239523.0
3                         9.6                   36.6          328239523.0
4                         9.6                   36.6          328239523.0

[5 rows x 33 columns]
```

Hem passat de tenir 35 columnes a tenir-ne 33. Per tant s'han eliminat les dues columnes esmentades.

4. Bibliografia

- Pàgina de descàrrega de la BBDD:

  https://www.kaggle.com/datasets/endofnight17j03/billionaires-statistics-dataset



Billionaires Statistics Dataset

Worldwide Wealth and Demographic Data: A Comprehensive Dataset

Data Card | Code (9) | Discussion (2) | Suggestions (0)

About Dataset

Certainly! Here's a description of each column:

Rank: The numerical ranking of a person or entity in a list or category.

finalWorth: The final worth or net worth of a person or entity, typically in terms of monetary value.

category: The category or classification of a person or entity, such as "entrepreneur", "investor", "celebrity", etc.

personName: The name of a person.

age: The age of a person.