**Certificate in Data Analytics for Business**

**March-May 2022**

**By: Marta Golec**

**Project Report**

*Title: Trends in Adoption of Children – Ireland and USA 1960-2019*

# **GitHub URL**

https://github.com/MartaGx/Adoption-Trends-IRL-vs-USA

# **Abstract**

This Project sets out to analyse the trends in adoption of children in Ireland and USA and to see if the high adoption rates in Ireland were influenced by the fact that a famous person (a.k.a. celebrity) adopted a child and received a large publicity and media coverage as a result.

# **Introduction**

I chose this particular topic, because I would like to see what the current trend is in terms of child adoption, amidst growing popularity of gestational surrogacy (a process where one person, who did not provide the egg used in conception, carries a foetus through pregnancy and gives birth to a baby for another person or couple)\*

# **Dataset**

I chose a variety of datasets and sources to show my ability to pull data from different sources: CSV, SQL, JSON, Web Scraping

**Dataset I – CSV**

Source:

https://www.aai.gov.ie/en/who-we-are/corporate-service/statistics.html

I have downloaded the data file, which was only available in pdf format, I converted it into Excel file, edited it (removed anything that was not an actual table) and saved as csv. I then imported it via pd.read\_csv command.

**Dataset II – JSON**

Source: https://data.gov.ie/organization/tusla?q=foster+care+general&sort=score+desc%2C+metadata\_modified+desc

I have filtered out the list of Datasets by first looking through what was available and then selecting the specific name of data set and putting it in the search box.

On the results lists it shows that both json and csv files were available for each Dataset, however when I was trying to download it, only two out of seven were available, the remainder were showing an error, so I downloaded them as csv and kept one as an example in order to show my ability to pull data from a json file.

**Dataset III – SQL**

Source:

https://www.insider.com/celebrities-adopted-kids-2018-3

Source:

conn=pymysql.connect(host='localhost', port=int(3306), user='root', passwd='passWORD\_2022', db='project\_certificate\_2022')

To show my ability to pull data from an SQL Server, I set up an account on MySQL Workbench portal and created a database with two tables in it (based on the data available on the Insider.com website), which are sharing a key column ‘name’. It was a very interesting experience and I also noticed how useful it is in working on SQL servers – every time I was changing something in the table, for example values or data types, it was immediately being reflected in the Jupyter notebook display.

**Dataset IV – Web Scraping**

Source:

https://www.johnstonsarchive.net/policy/adoptionstats.html

After doing some research online looking for adoption statistics in the USA, I came across this webpage titled “Historical statistics on adoption in the United States, plus statistics on child population and welfare” created by Robert Johnston, which is the most comprehensive dataset in my whole project.

# **Implementation Process**

**Dataset I – CSV**

After importing the Dataset, I have prepared the data for analysis by creating a DataFrame. dropping duplicates and dropping missing values. Then I went on to renaming column headings and converting data types – in order to be able to carry out operations on numbers. I computed min and max values of the Dataset and sliced the DF in order to create a subset. I made two line plots using Matplotlib.

**Dataset II – JSON**

I imported the necessary package and loaded the file. Using FOR loop, I organised the data in a key : value string. I examined the ‘soup’ and noticed that the file came with datatype information that I didn't know how to normalize (or if they were even needed), so I just copied the 'features:' part of the file and created a list of dictionaries, so that I could convert them into a DataFrame. I sliced the DataFrame, as only the data available at the end of the year 2021 are needed (reading Nov column, as Dec was null). I was trying to slice the columns by -2 index position using this command: ch\_fc\_2021.iloc[: , [-2]], but I kept getting an error message: "positional indexers are out-of-bounds”. The -1 position was working, so I dropped the Dec column and that way was able to select Nov by calling it -1 position. I was able to turn the ‘soup’ into a table by using the command json\_normalize.

# I imported the rest of the files as csv, because many of the files available in json had an error in source. I also noticed that importing csv files is much quicker.

# I sliced the DataFrame from Dataset I to merge with DataFrame from Dataset II. I then merged both DataFrames on Right Join, as the right DataFrame had less rows (years), so they needed to be joined only on the years where data existed. I created a new column to show the percentage and then used the .mean() function to calculate the average of that percentage – to use in this report. I created an overlay of two barplots to visualise the yearly number of adopted children against the total yearly number of children in Foster Care.

**Dataset III – SQL**

# I set up a connection with MySQL database using pymysql package. I retrieved a table and converted it into a new DataFrame, which I then cleaned. I used Numpy to summarize numerical data- min, max, mean and median.

# I creating two bar subplots and a histogram with custom colours.

# I pulled a second table from MySQL, which shows the amount of followers each celebrity has on Instagram. I chose to do Nicole Kidman & Tom Cruise jointly, as they were a couple back when they adopted and they were both very popular actors. I merged both DataFrames on the Name key – I chose Inner Join, because I wanted to see only the results, where data existed. I sorted values from the highest to the lowest to see how popular the adopting celebrity was. I wanted to create a new column showing the level of popularity. I chose 3 tiers: high, medium and low. In order to calculate the level of popularity based on the amount of followers on Instagram I created a custom function, which was a Conditional Statement and it populated the column itself according to the given ranges.

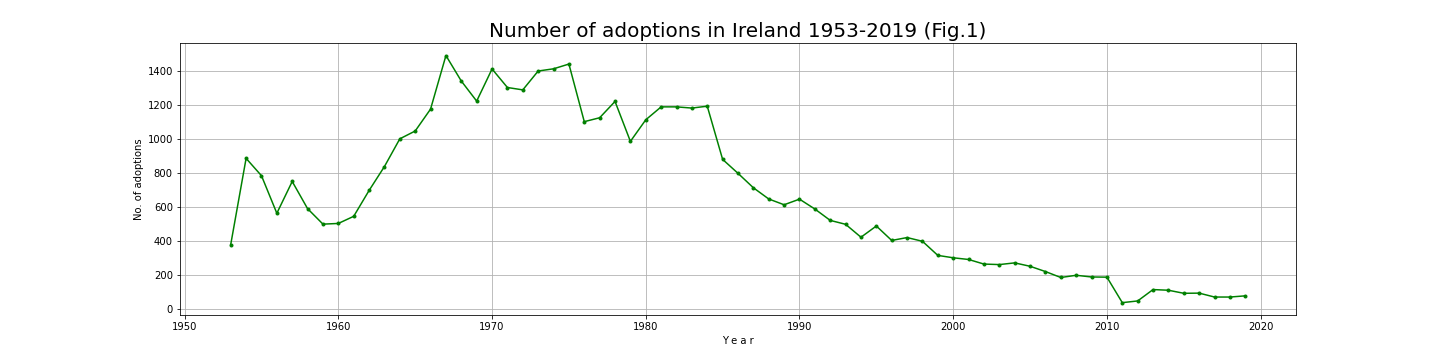
# I then merged the two DataFrames to obtain the year of adoption and how popular the adopting celebrity currently is to show where, potentially, they could have influenced the adult population to adopt a child. I then visualised the concept using a barplot and a scatterplot.

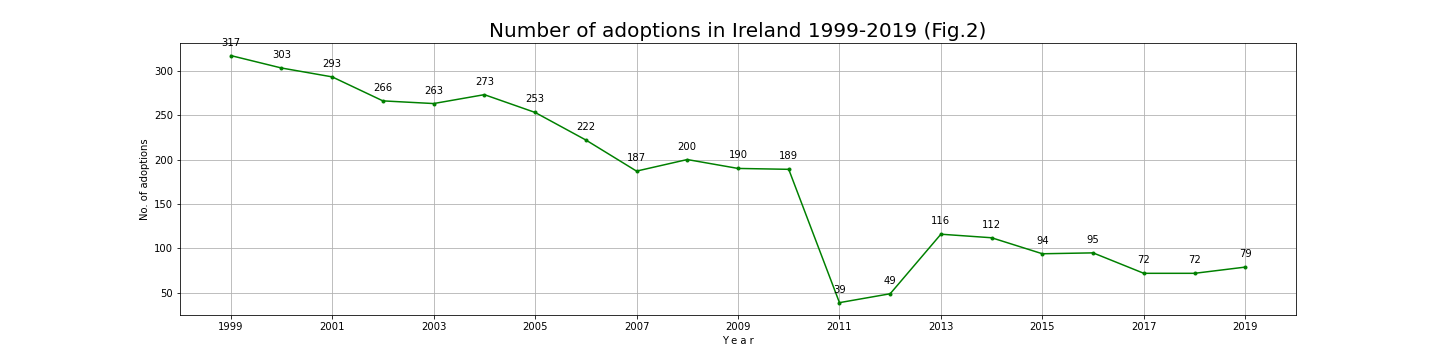
**Dataset IV – Web Scraping**

# I imported the necessary packages and pulled the data via url. I changined html to Python-friendly format via lxml parser. I obtained the ‘soup’, which is all the information that is on that webpage. I found the table tag <table border="2" cellpadding="5">, which was not a very straightforward tag, because initially I thought I have to look for <table> tag on its own and the tag<id> next to it. But I eventually managed to obtain information from that tag. I merged DataFrame from Dataset I with the new DataFrame, dropped null rows and ploted two lineplots to show how, relatively to the USA, the adoption trend has significantly dropped in Ireland and to see if there is any correlation between celebrity adoption and adoption trend in the USA.

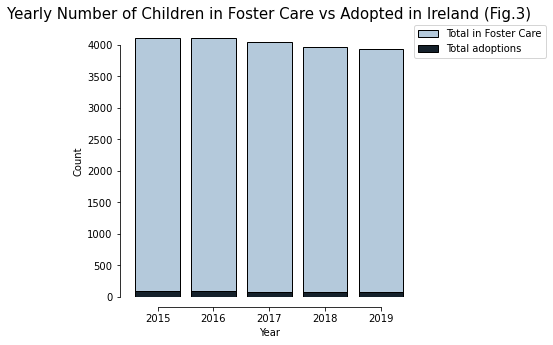
# **Results**

In the 1960’s and 70’s the adoption in Ireland was relatively high, with the peak reached in 1968 when **1493** children were adopted (Fig.1). There has been a continuous dropbetween the years 1984 and 2011 **(97%)** - the lowest number was in 2011, when only **39** children were adopted (Fig.2). Since 2017, the yearly adoption rate oscillates between 70 and 80.

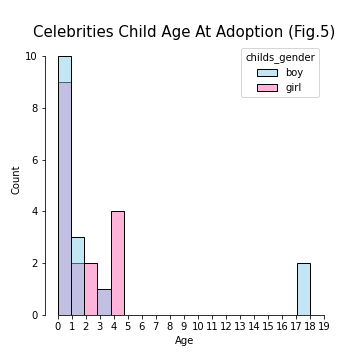


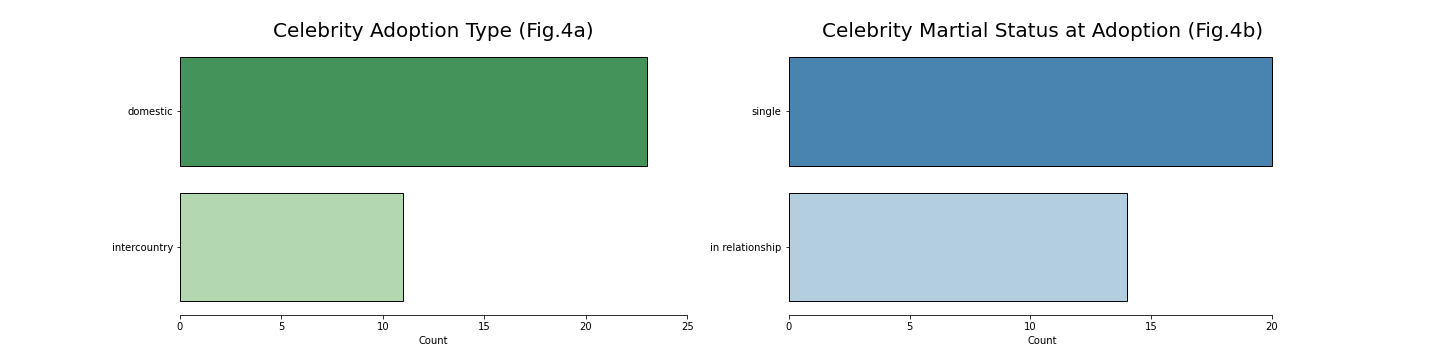


There are around **4,000** children in Foster Care each year (data from years 2015 – 2019). On average, only **2%** of this group is being adopted (Fig.3)

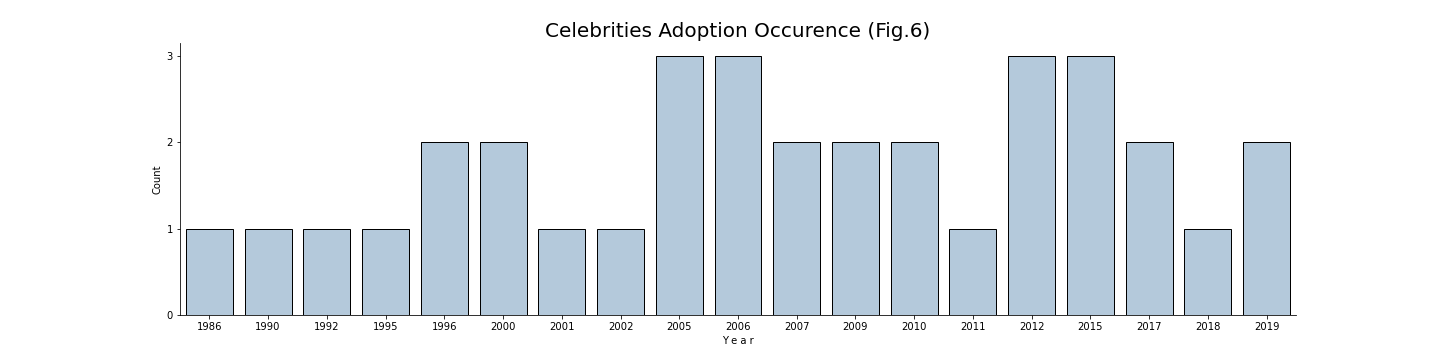


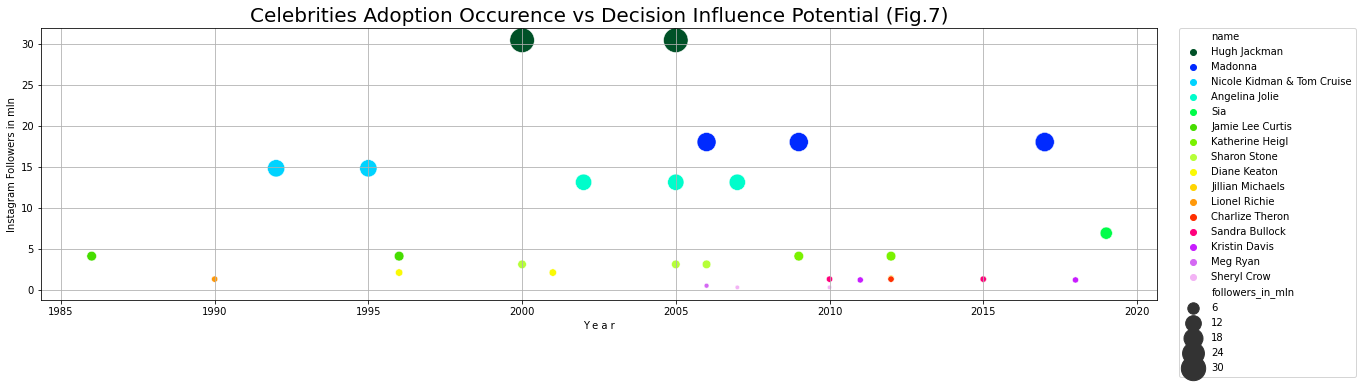
Statistics show that among the famous people who adopted a child/ children, these adoptive parents seek to adopt children right after they were born (Fig.5), which indicates that the process of adoption starts when the biological mother is still pregnant. There is a huge gap from the age of 4 upwards, suggesting that Celebrities do not want to adopt children older than this. The only Celebrity (popularity measured based on the amount of followers on Instagram) who has adopted above this age was Sia. She became a mother to two 18-year-old boys, who grew out of Foster Care.





There were twice as many domestic adoptions among the Celebrities as intercountry. (Fig.4a). An interesting fact is that among Celebrities, more children were adopted by single parents than couples. (Fig.4b)





In years 2005 and 2006 (Fig.6), there was a higher number of children adopted by top 3 most popular Celebrities - out of all the Celebrities who adopted: Hugh Jackman (30mln followers on Instagram currently), Madonna (18mln) and Angelina Jolie (13mln) (Fig.7), which resulted in global publicity and extensive media coverage. If adoption process takes on average 1 year\*\*, it correlates with the little spike in adoptions in the USA in 2007 – **153,000**, preceded by 131,000 in 2006 and followed by 139,000 in 2008 (Fig.8). This occurrence had no influence on the adoption trend in Ireland.

Chart, line chart

Description automatically generated

# **Insights**

* The fact that famous people (mostly from USA) have adopted children and have received media coverage as a result, did not help the child adoption rate in Ireland, which is **93%** less now than in mid-1980’s.
* The same fact has had a positive correlation with the increase of adoptions in the USA - **17%** year on year 2006 – 2007
* The children adopted by celebrities are mostly adopted just after their birth (**56%**), there are almost no adoptions of children aged 5 and up.
* There are currently around **4,000** children in State care in Ireland, some coming from disfunctional families who were in temporary foster care, but only **2%** of these children has the opportunity to live with a new family.
* Speaking more about the adoption in the public space could encourage more people to adopt children.

# **Machine Learning**

# In this particular subject, adoption, machine learning could be used to analyse the amount of media coverage (tabloid and press articles, social media activity) needed to design a national awareness campaign encouraging people to adopt children. I would use regression methods to predict trends.

# **References**

\* Department Of Health (2021). Gestational Surrogacy Fact Sheet [online]. Available from:

<https://health.ny.gov/community/pregnancy/surrogacy/gestational_surrogacy_fact_sheet.htm> [accessed 19 May 2022]

\*\* National Adoption Foundation (2022). How Long Does It Take To Adopt A Child [online]. Available from: <https://fundyouradoption.org/resources/how-long-does-it-take-to-adopt-a-child/> [accessed 24 May 2022]