INVESTIGATION INTO WHETHER BABY NAMES IN THE UK ARE INFLUENCED BY MOVIE RELEASES

1

CODE FIRST GIRLS

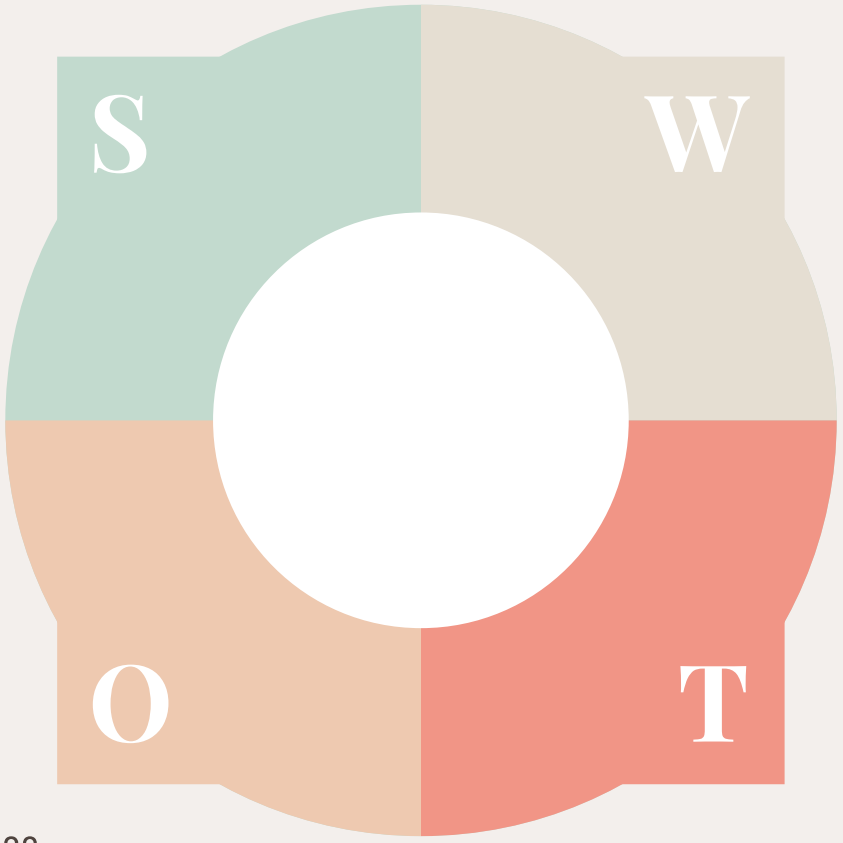
SUE SCOTT
GEORGIA GRANT
FLAVIA FONSECA
KATIE PAPWORTH
LAURA FERNANDES

INTRODUCTION

SWOT ANALYSIS

TRENGTHS

matics/statistics
hysics
ical design
ical skills
t management
e management
vity
unication



WEAKNE

- Coding ex
- Coding pr
- Work com
- Too many
- Not enoug consolida topics

ORTUNITIES

from each other
ng in a group to produce
oject

THREAT

- Covid (illn
- Remote w communi ta (API)
- Working with live data

PROJECT TASKS DISTRIBUTION

TASK	MEMBERS INVOLVED
DATA COLLECTION AND RESEARCH	ALL
DATA CLEANING	ALL
API DATA	LAURA, SUE GEORGIA
GRAPHS AND ANALYSIS	KATIE, LAURA
SQL DATABASE	FLAVIA, LAURA
MODELLING, PREDICTIONS, HYPOTHESIS TESTING	GEORGIA, KATIE
REPORT WRITING	ALL
TEAM MANAGEMENT	SUE

TOOLS AND LIBRARIES USED

SCIENTIFIC LIBRARIES	• Pandas • NumPy • SciPy • SciKit-Learn
VISUALISATION LIBRARIES	• MatPlotLib • Seaborn
MODULES	• Scipy.stats • sklearn.linear_model • Requests • Openpyxl
KEY FUNCTIONS	• LinearRegression • Wilcoxon (statistical test)
OTHER TOOLS	• SQL Workbench • Slack & WhatsApp • Zoom - meetings • Trello - project management tool

DATA COLLECTION

DATA SOURCES

Office for National Statistics

THE MOVIE DB

DATA CLEANING

Contents

Names for baby boys in England and Wales[1], 1996 to 2

	2020		2019	
Name	**Rank**	**Count**	**Rank**	**Count**
A	3,848	4	2,104	10
A:JAY	2,960	6	:	:
A.J.	:	:	:	:

Contents

Names for baby girls in England and Wales[1], 1996 to 2

	2020		2019	
Name	**Rank**	**Count**	**Rank**	**Count**
A'IDAH	:	:	4,686	4
A'ISHA	:	:	2,458	10
A'ISHAH	:	:	:	:

```
Out[5]: [{'adult': False,
  'backdrop_path': '/5RuR7GhOI5fElADXZb0X2sr9w5n.jpg',
  'genre_ids': [16, 35, 10751, 14],
  'id': 568124,
  'original_language': 'en',
  'original_title': 'Encanto',
  'overview': "The tale of an extraordinary family, the Madriga
ical house, in a vibrant town, in a wondrous, charmed place cal
child in the family with a unique gift from super strength to t
she discovers that the magic surrounding the Encanto is in dang
might just be her exceptional family's last hope.",
  'popularity': 2595.206,
  'poster_path': '/4j0PNHkMr5ax3IA8tjtxcmPU3QT.jpg',
  'release_date': '2021-11-24',
  'title': 'Encanto',
  'video': False,
  'vote_average': 7.3,
  'vote_count': 344},
 {'adult': False,
```

	Name	2020 Rank	2019 Rank
4	A	3848	2104
5	A:JAY	2960	4813
7	AABAN	3848	3937
8	AABID	4608	3937
9	AADAM	456	327
12	AADEN	4813	4813

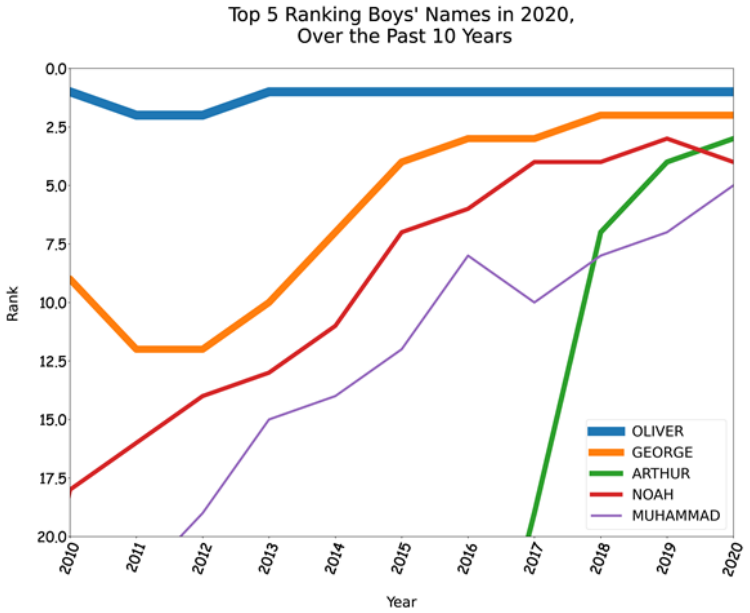
	Name	2020 Rank	2019 Rank
5	A'ISHA	5877	2458
6	A'ISHAH	5877	5877
13	AADHYA	969	1252
19	AADYA	2042	1497
20	AAEESHA	5877	5877
21	AAFIA	2843	4001

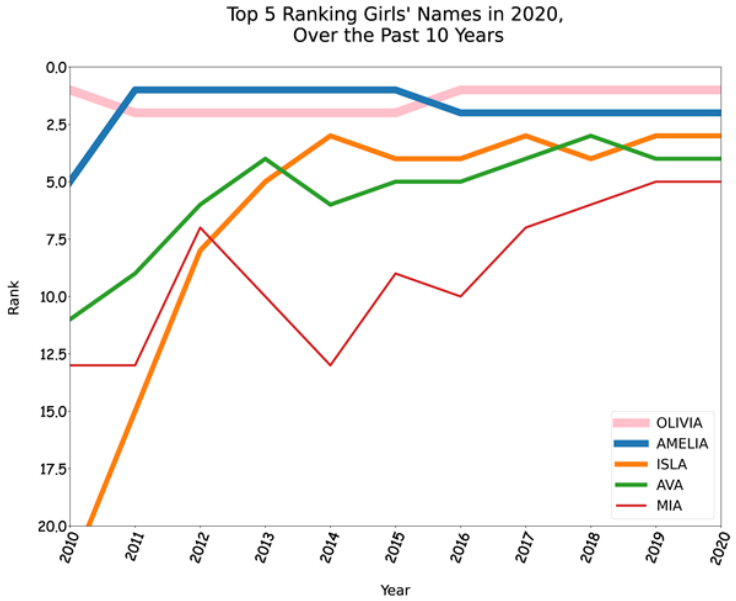
	title	rating	release_year
117	Up	7.9	2009
91	Coraline	7.8	2009
161	The Princess and the Frog	7.1	2009

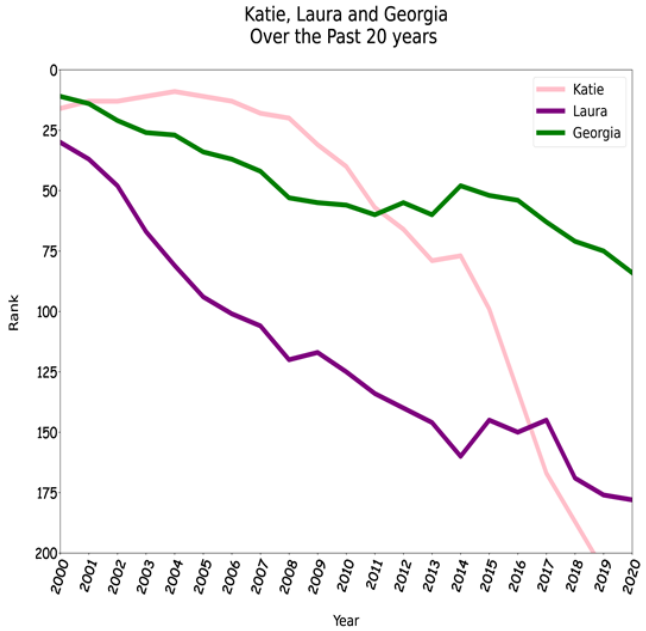
	title	rating	release_year
216	La leyenda de la Llorona	7.7	2011
160	Barbie: Princess Charm School	7.4	2011
188	Soul Surfer	7.1	2011

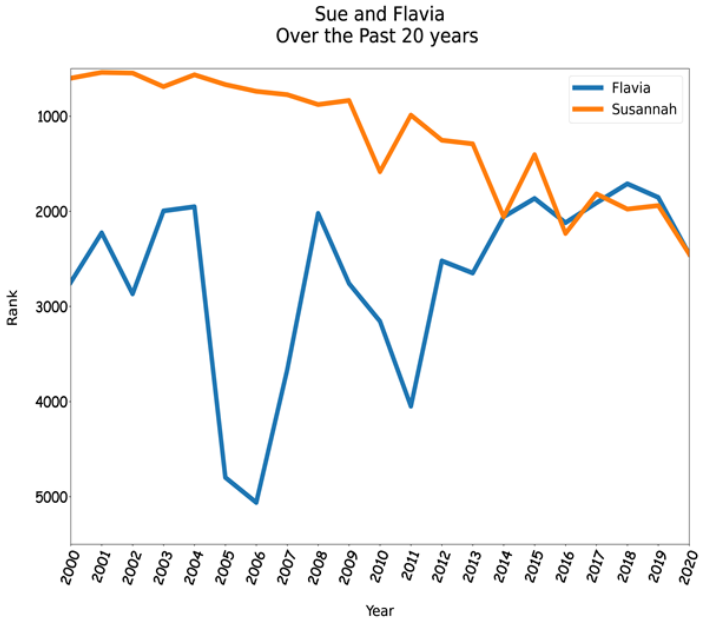
	title	rating	release_year
77	Frozen	7.3	2013
125	Monsters University	7.0	2013
109	The Croods	6.9	2013

EXPLORING THE DATA

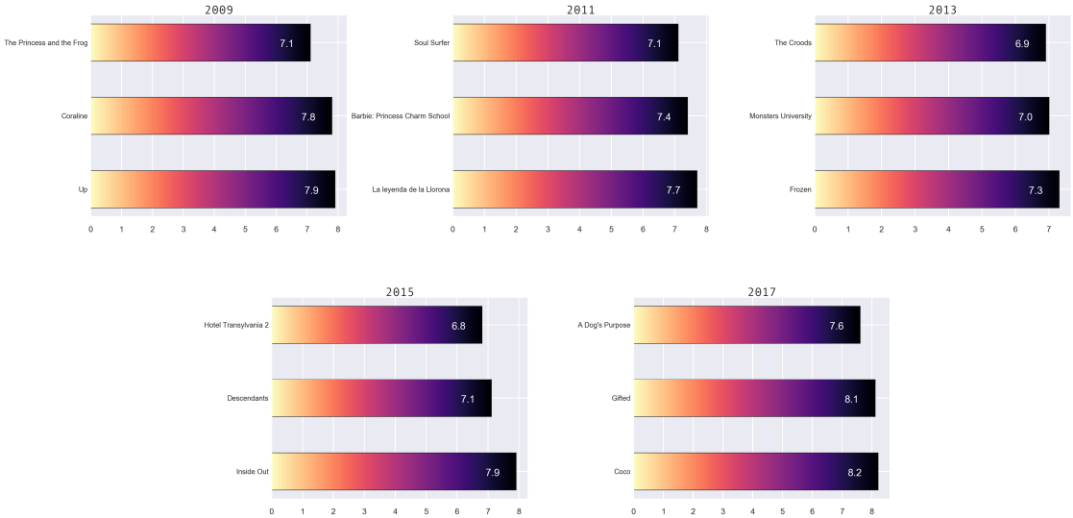




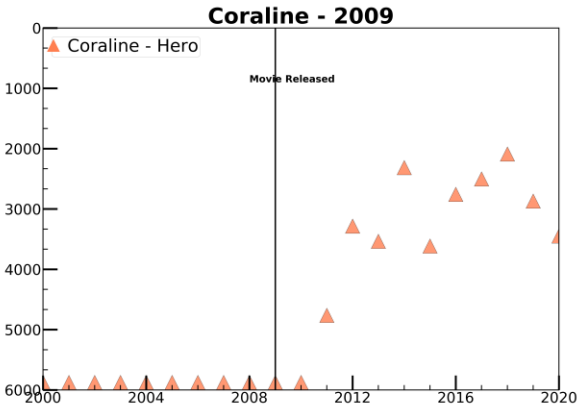




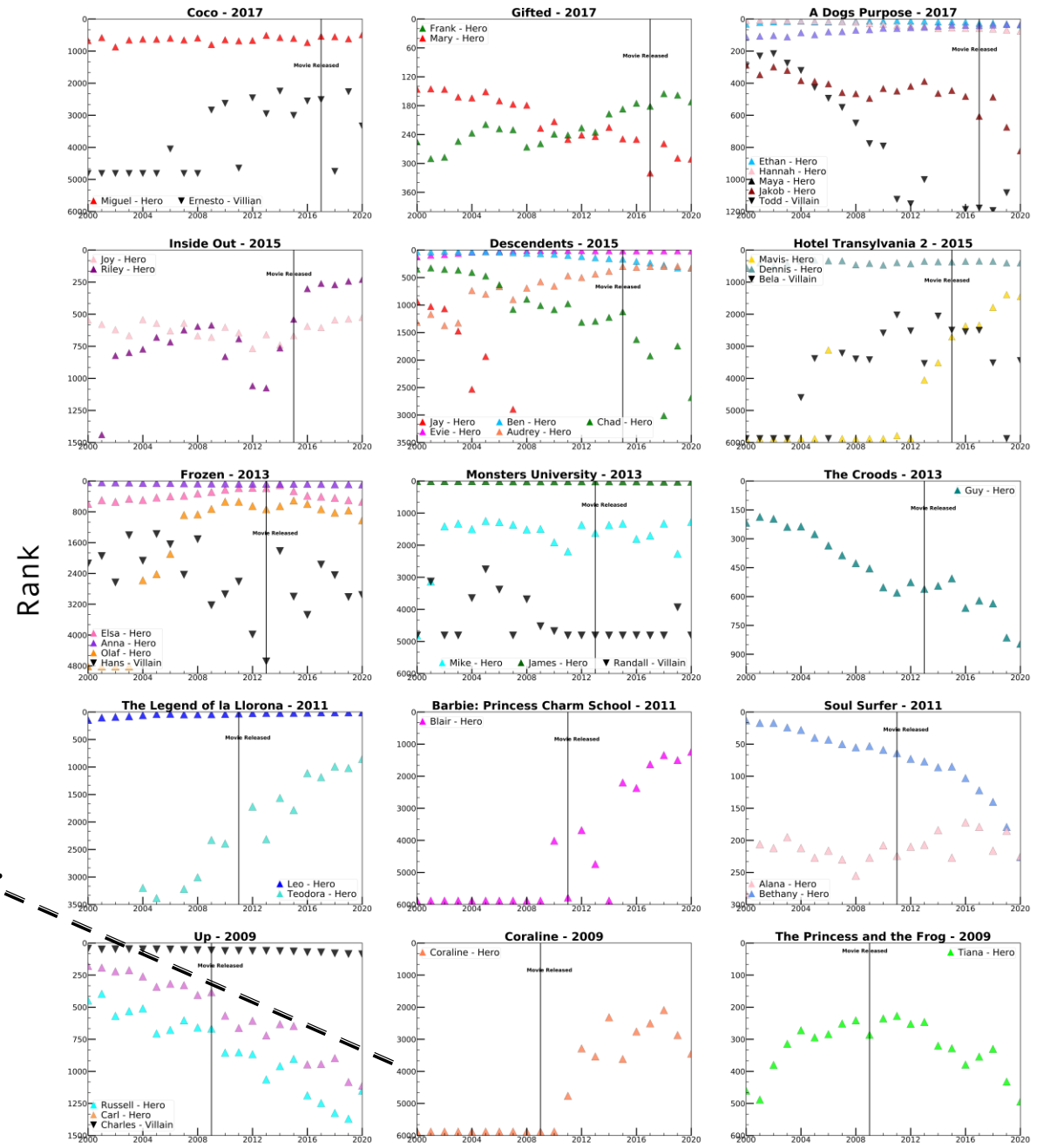
Top 3 Most Popular Movies in 2009, 2011, 2013, 2015 and 2017



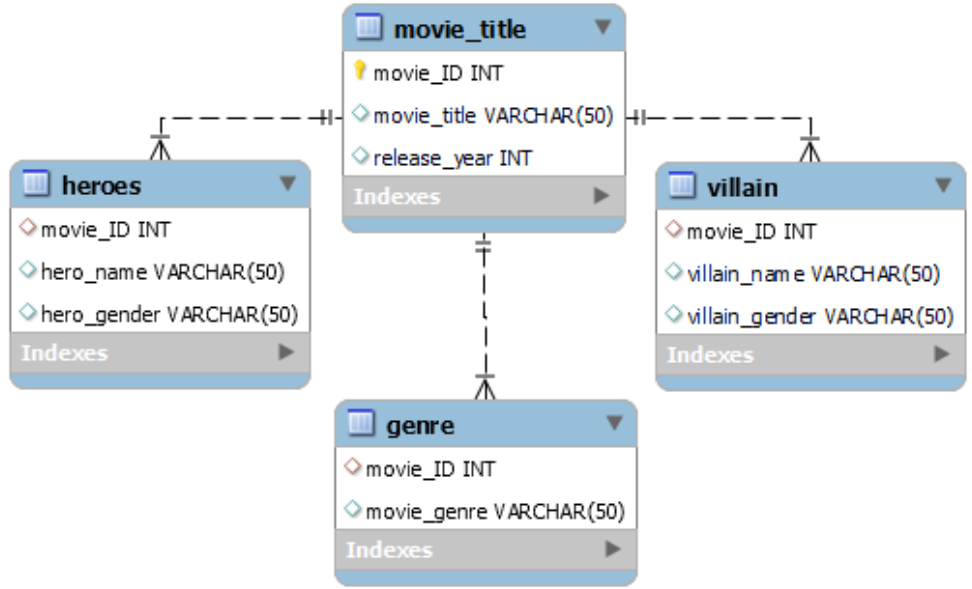
EXPLORING THE DATA



The Influence of Movie Heros and Villains on Baby Names



DATABASE CREATED




```
hero_characters_2009 = 'SELECT h.movie_ID, h.hero_name, h. hero_gender, m. movie_title FROM heroes h RIGHT JOIN movie_title m
df_3h = pd.read_sql_query(hero_characters_2009, connection)
df_3h
```

:

	movie_ID	hero_name	hero_gender	movie_title
0	13	Russell	m	Up
1	14	Coraline	f	Coraline
2	15	Tiana	f	The Princess and the Frog

KEY STATISTICS SPREADSHEET

FUNCTION TO PULL KEY STATS FROM CLEAN BABY NAME DATA:

```
def girl_2013_stats(name):
        rank2012 = girls_rank_df_t[name][8]
        rank2013 = girls_rank_df_t[name][7]
        rank2014 = girls_rank_df_t[name][6]
        print(name, "STATS:\n")
        print("2012 Rank:", rank2012)
        print("2013 Rank:", rank2013)
        print("2014 Rank:", rank2014)


        pc_change_2012_to_2013 = (rank2013 - rank2012)*100/rank2012
        print("\n2012 to 2013 percentage change:", np.around(-pc_change_2012_to_2013, decimals=2), "%")

        pc_change_2013_to_2014 = (rank2014 - rank2013)*100/rank2013
        print("2013 to 2014 percentage change:", np.around(-pc_change_2013_to_2014, decimals=2), "%")

        mean_3yrs_before = np.around(np.mean(girls_rank_df_t[name][8:11]), decimals=2)
        print("\nMean rank 3 years before release:", mean_3yrs_before)
        mean_3yrs_after = np.around(np.mean(girls_rank_df_t[name][6:9]), decimals=2)
        print("Mean rank 3 years after release:", mean_3yrs_after, "\n\n")
```

```
girl_2013_stats('ELSA')
```

```
ELSA STATS:

2012 Rank: 177.0
2013 Rank: 183.0
2014 Rank: 104.0

2012 to 2013 percentage change: -3.39 %
2013 to 2014 percentage change: 43.17 %

Mean rank 3 years before release: 194.33
Mean rank 3 years after release: 154.67
```

SPREADSHEET WITH HERO/VILLAIN NAME KEY STATS:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	Movie	Rating	Name	Rank in year before release	Rank in year of release	Rank in year after release	Pc change year before to year of release	Pc change year of to year after release	Mean rank 3 years before release	Mean rank 3 years after release	Pc change mean rank before to mean rank after
2	2009	The Princess and the Frog	7.1	TIANA	241	286	235	-18.67	17.83	258.67	238	7.99
3	2009	Up	7.9	RUSSELL	658	667	854	-1.37	-28.04	646	857.33	-32.71
4	2009	Coraline	7.8	CORALINE	5877	5877	5877	0	0	5877	4642	21.01
5	2011	The Legend of Llorona	7.7	LEO	36	27	20	25	19.44	38.67	17.67	54.31
6	2011	Soul Surfer	7.1	ALANA	208	224	210	-7.69	6.73	230	200.33	12.90
7	2011	Barbie: Princess Charm School	7.4	BLAIR	4012	5785	3681	-44.19	52.44	5255.33	4765.67	9.32
8	2013	Frozen	7.3	ELSA	177	183	104	-3.39	43.17	194.33	154.67	20.41
9	2013	Monsters University	7	MIKE	1367	1617	1370	-18.29	15.28	1824	1451.33	20.43
10	2013	The Croods	6.9	GUY	526	560	544	-6.46	2.86	552.67	543.33	1.69
11	2015	Inside Out	7.9	RILEY	763	539	301	29.36	44.16	965.33	277	71.31
12	2015	Inside Out	7.9	JOY	740	666	594	10	10.81	722.33	581	19.57
13	2015	Descendents	7.1	EVIE	14	15	16	-7.14	-6.67	13	15.33	-17.92
14	2015	Hotel Transylvania	6.8	MAVIS	3514	2705	2367	23.02	12.5	4480	2165.33	51.67
15	2015	Hotel Transylvania	6.8	DENNIS	365	368	350	-0.82	4.89	379.33	347.33	8.44
16	2017	A Dog's Purpose	7.6	ETHAN	19	26	25	-36.84	3.85	18.33	30.33	-65.47
17	2017	Coco	8.2	MIGUEL	732	537	549	26.64	-2.23	635.67	551	13.32
18	2017	Gifted	8.1	MARY	250	320	259	-28	19.06	241.33	279.67	-15.89
19												
20												

Heroes | Villains | All | +

WHAT WE FOUND

INFLUENTIAL MOVIES

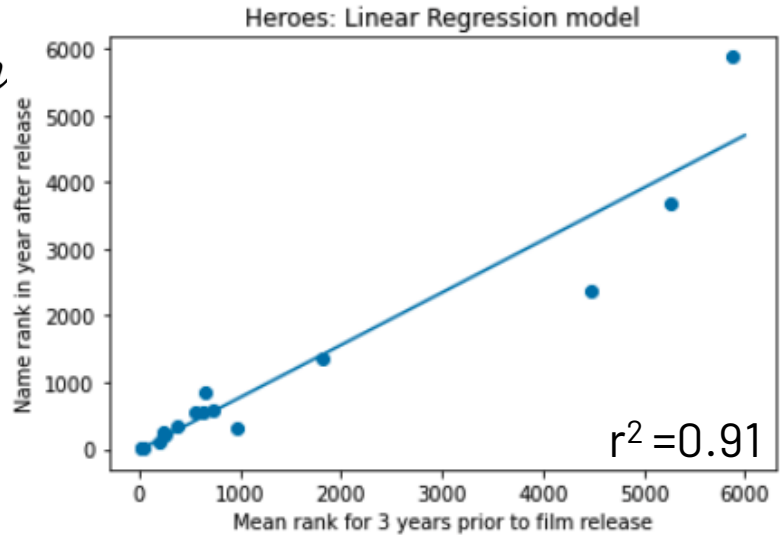
MOVIE	NAME	% CHANGE IN RANK
INSIDE OUT	RILEY	71.3%
THE LEGEND OF LLORONA	LEO	54.3%
HOTEL TRANSYLVANIA	MAVIS	51.7%
CORALINE	CORALINE	21.0%
MONSTERS UNIVERSITY	MIKE	20.4%
FROZEN	ELSA	20.4%
INSIDE OUT	JOY	19.6%

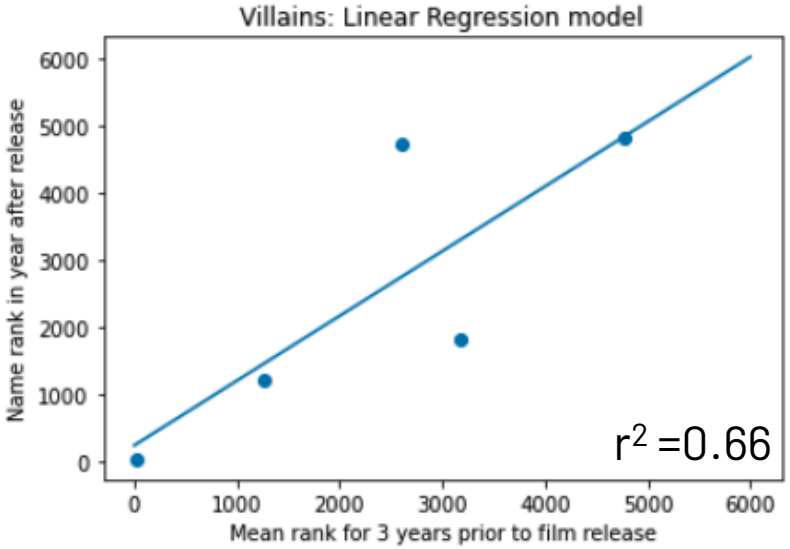
WHAT WE FOUND

HEROES VS VILLAINS

STAT	HEROES	VILLAINS
MEAN RANK 3 YRS BEFORE RELEASE	1313	2367
MEAN RANK 3 YRS AFTER RELEASE	1006	2609
MEAN % CHANGE IN AVG RANK	+10.6%	-19.3%

PREDICTIVE MODELLING





2021 BABY NAME PREDICTIONS

NAME	2020 MOVIE	2020 RANK	2021 PREDICTED RANK
LAUREL	ONWARD	1507	1121
ELEANOR	GODMOTHERED	53	31
TIMOTHY	THE WILLOUGHBYS	388	280
MAXWELL	WONDER WOMAN 1984	157	376

```
Difference_Before_After_Hero = hero_data['Mean rank 3 years before release'] - hero_data['Mean rank 3 years after release']
Difference_Before_After_Hero
```

HYPOTHESIS TESTING

HEROES

Test: Rank unchanged or decreases?

Result: Strong evidence to suggest that the rank of a baby name becomes more popular

VILLAINS

Test: Rank unchanged or increases?

Result: Weaker evidence to suggest that the rank of a baby name becomes less popular

```
d1a = Difference_Before_After_Hero
w, p = wilcoxon(d1a, alternative = 'greater')
w, p
```

```
(129.0, 0.0054931640625)
```

```
d2a = Difference_Before_After_Villain
w, p = wilcoxon(d2a, alternative = 'less')
w, p
```

```
(1.0, 0.0625)
```

CONCLUSION

1. THERE WAS A CHANGE IN RANK AFTER THE MOVIES INVESTIGATED WERE RELEASED
2. HEROES NAMES: INCREASE IN POPULARITY BY 10.6%
3. VILLAIN NAMES: DECREASE IN POPULARITY BY 19.3%
4. MOST INFLUENTIAL MOVIE: INSIDE OUT (*RILEY*'S POPULARITY INCREASED BY 71.3%)


CODE
FIRST
GIRLS
SUE SCOTT
GEORGIA GRANT
FLAVIA FONSECA
KATIE PAPWORTH
LAURA FERNANDES

QUESTIONS

Slide Order

Slide 1: Title page

Slide 2: Introduction/SWOT analysis/ task distribution SUE

Slide 3: Tools and Libraries used SUE

Slide 4: Data Collection SUE

Slide 5: Data Cleaning SUE

Slide 6: Exploring the Data SUE

Slide 7: Exploring the Data LAURA

Slide 8 :Database Created FLAVIA

Slide 9: Key Statistics Spreadsheet KATIE

Slide 10: Influential Movies KATIE

Slide 11: Heroes vs Villains KATIE

Slide 12: Predictive modelling KATIE

Slide 13:Hypothesis Testing GEORGIA

Slide 14: Conclusion GEORGIA

Slide 15: Any questions?