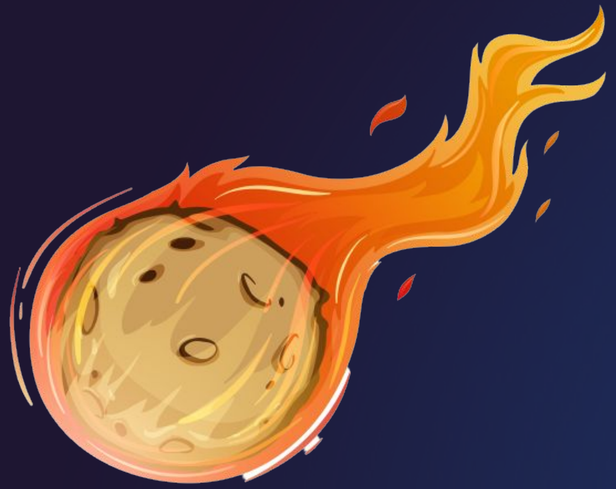
Meteorite Landings

Data Science Project

by Avihoo Elias & Aviv Zano

What are Meteorites?

Meteors are small pieces of space debris that fall down on the earth, The vast majority of meteors burn up high in the atmosphere, but those few that reach the ground are called meteorites. The meteorites differs in mass, composition and location.



Our Research Questions

- ❏ Can we predict the year of landing based on meteorites coordinates?
- ❏ Are Meteorites more likely to be found in specific areas in the world?

Data Science Phases

01

Data Acquisition

02

Data Cleaning

03

EDA

04

Machine Learning

01

Data Acquisition

Data Source - NASA

- We used sodapy to retrieve the data from an API provided by NASA:

  https://data.nasa.gov/Space-Science/Meteorite-Landings/gh4g-9sfh

Data Source - NASA

```
results_df.head()
```

	name	id	nametype	recclass	mass	fall	year	reclat	reclong	geolocation	:@computed_region_cbhk_fwbd	:@computed_region
0	Aachen	1	Valid	L5	21	Fell	1880-01-01T00:00:00.000	50.775000	6.083330	{'latitude': '50.775', 'longitude': '6.08333'}	NaN	
1	Aarhus	2	Valid	H6	720	Fell	1951-01-01T00:00:00.000	56.183330	10.233330	{'latitude': '56.18333', 'longitude': '10.23333'}	NaN	
2	Abee	6	Valid	EH4	107000	Fell	1952-01-01T00:00:00.000	54.216670	-113.000000	{'latitude': '54.21667', 'longitude': '-113.0'}	NaN	
3	Acapulco	10	Valid	Acapulcoite	1914	Fell	1976-01-01T00:00:00.000	16.883330	-99.900000	{'latitude': '16.88333', 'longitude': '-99.9'}	NaN	
4	Achiras	370	Valid	L6	780	Fell	1902-01-01T00:00:00.000	-33.166670	-64.950000	{'latitude': '-33.16667', 'longitude': '-64.95'}	NaN	

02

Data Cleaning

Data Cleaning

- We removed the “@computed” columns - they were full of NaN values.
- We dropped all of the rows that contained null cells (about 7,000 out of 45,000)
- We recreated our geolocation column - replaced the json format with a tuple format.

```
# Data Cleaning

results_df.drop(labels=['geolocation',':@computed_region_cbhk_fwbd',':@computed_region_nnqa_25f4'],axis=1,inplace=True)
results_df.dropna(inplace=True)
```

```
results_df["geolocation"] = list(zip(results_df.reclat, results_df.reclong))
```

Dataset

```
results_df.head()
```

	name	id	nametype	recclass	mass	fall	year	reclat	reclong	geolocation
0	Aachen	1	Valid	L5	21	Fell	1880-01-01T00:00:00.000	50.775000	6.083330	(50.775000, 6.083330)
1	Aarhus	2	Valid	H6	720	Fell	1951-01-01T00:00:00.000	56.183330	10.233330	(56.183330, 10.233330)
2	Abee	6	Valid	EH4	107000	Fell	1952-01-01T00:00:00.000			000)
3	Acapulco	10	Valid	Acapulcoite	1914	Fell	1976-01-01T00:00:00.000			000)
4	Achiras	370	Valid	L6	780	Fell	1902-01-01T00:00:00.000			000)

```
results_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 38115 entries, 0 to 45715
Data columns (total 10 columns):
 #   Column       Non-Null Count  Dtype 
---  ------       --------------  ----- 
 0   name         38115 non-null  object
 1   id           38115 non-null  object
 2   nametype     38115 non-null  object
 3   recclass     38115 non-null  object
 4   mass         38115 non-null  object
 5   fall         38115 non-null  object
 6   year         38115 non-null  object
 7   reclat       38115 non-null  object
 8   reclong      38115 non-null  object
 9   geolocation  38115 non-null  object
dtypes: object(10)
memory usage: 3.2+ MB
```

Dataset - Columns

- **name:** The name of the meteorite
- **id:** a unique identifier of each meteorite
- **nametype:**
  - Valid: a typical meteorite
  - Relict: a meteorite that has been highly degraded by weather on Earth
- **recclass:** the class of the meteorite
- **mass:** the mass of the meteorite (in grams)
- **fall:**
  - Fell: the meteorite's fall was observed
  - Found: the meteorite's fall was not observed
- **year:** the year the meteorite fell, or the year it was found
- **reclat:** the latitude of the meteorite's landing
- **reclong:** the longitude of the meteorite's landing

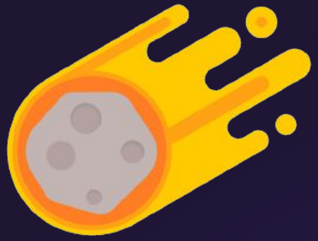
03

EDA

Visualizing and analyzing the data

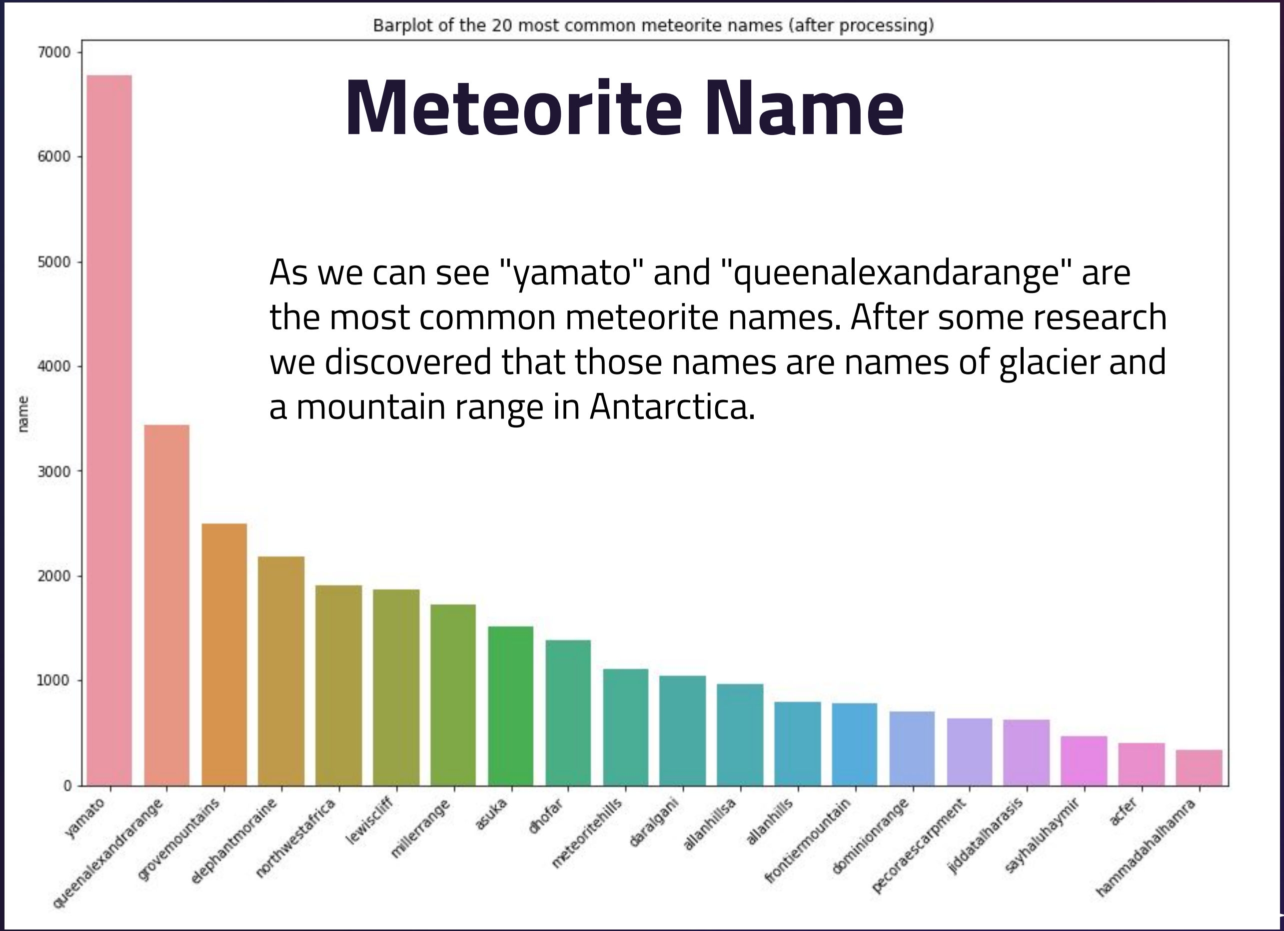
Features Analysis

As part of the feature analysis we are going to analyze each feature (column) to better understand our data.

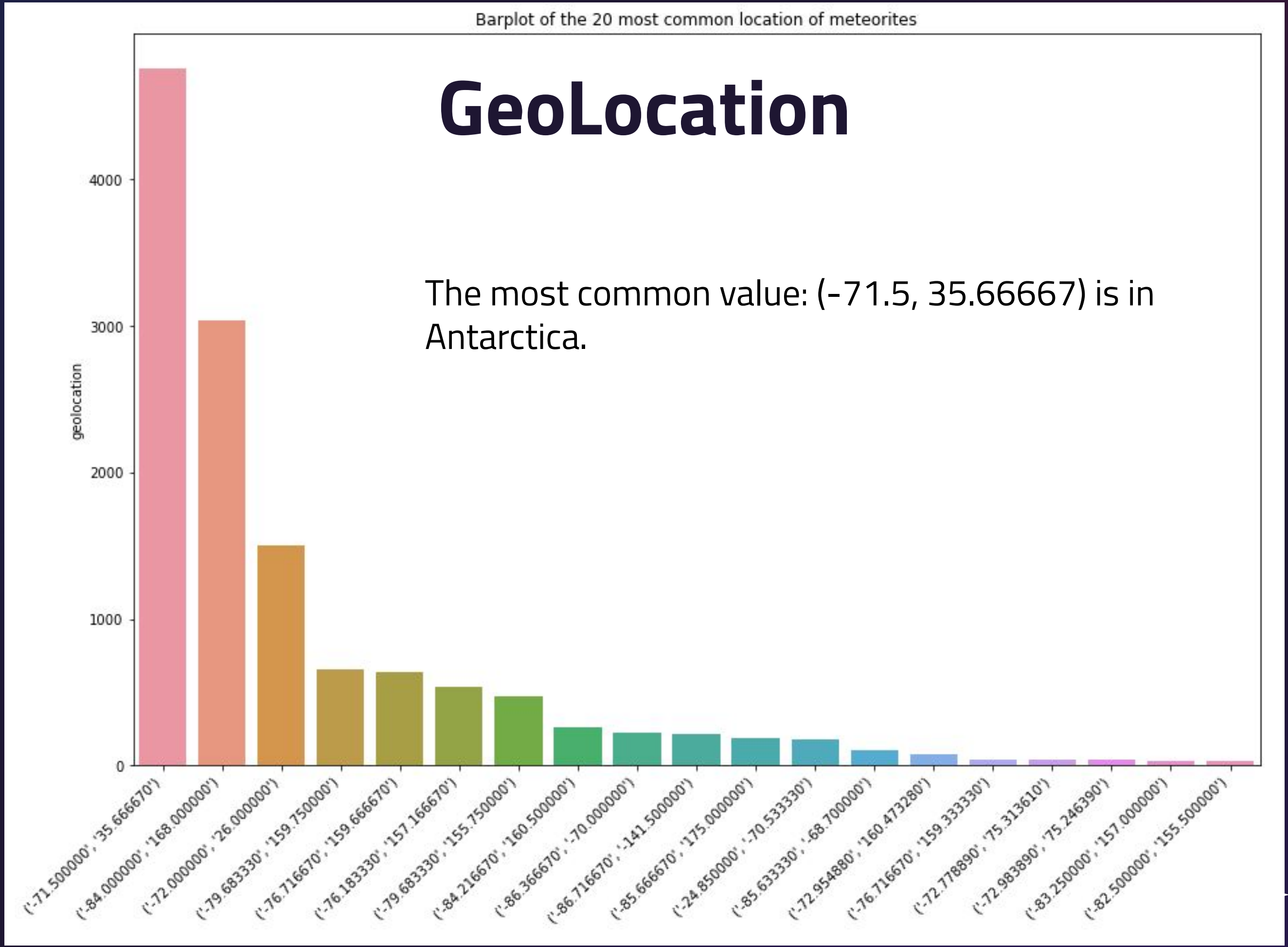


Meteorite Name

As we can see "yamato" and "queenalexandarange" are the most common meteorite names. After some research we discovered that those names are names of glacier and a mountain range in Antarctica.



GeoLocation



The most common value: (-71.5, 35.66667) is in Antarctica.

GeoLocation

GeoLocation

Antarctica is the best place to find meteorites. Its dry, cold climate perfectly preserves the asteroid fragments that make it through the atmosphere and crash on Earth.

Recclass

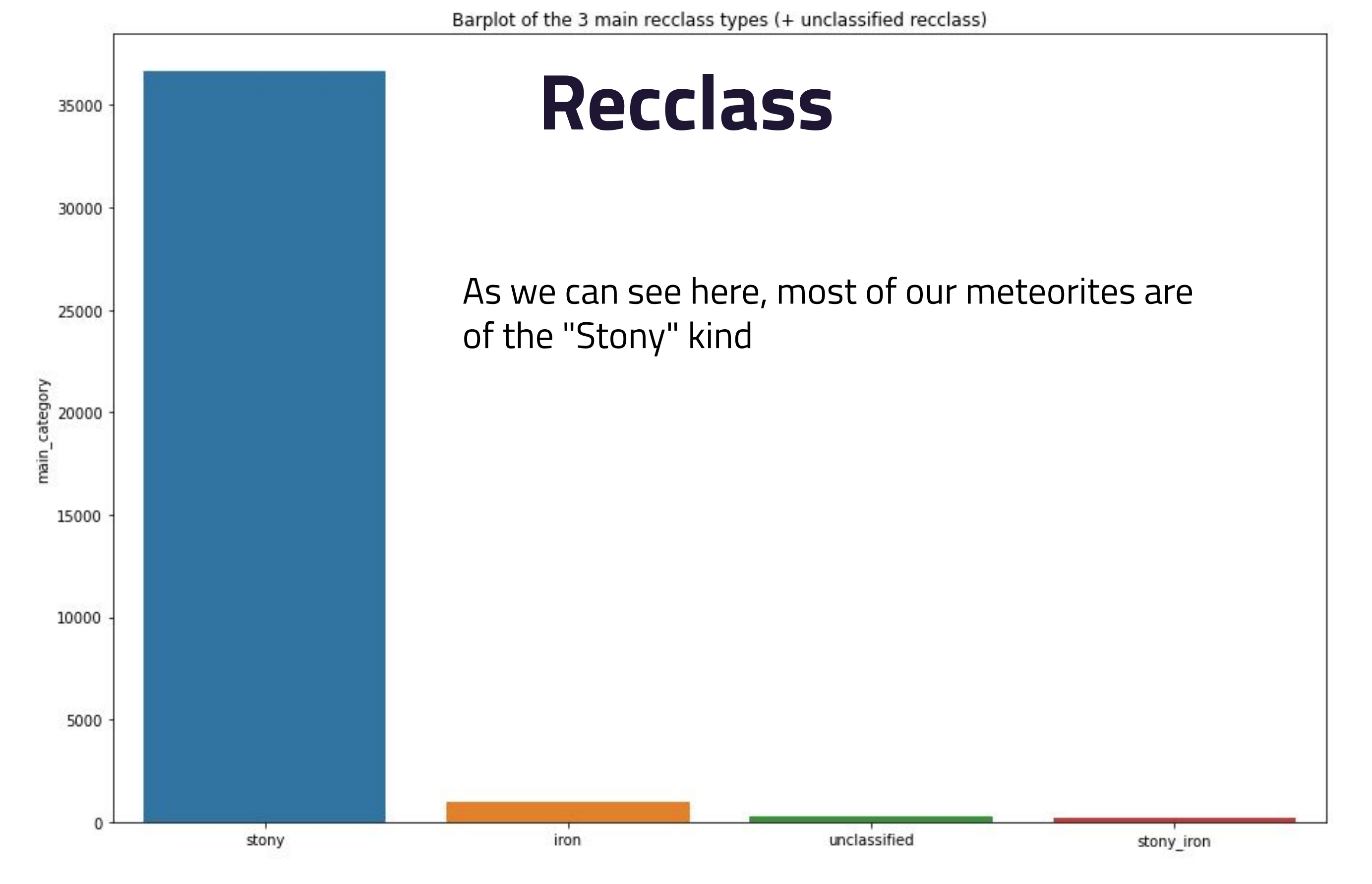
This column describes the class of the meteorite.

There are a lot of different classes of meteorites, how ever we can divide them to 3 main categories as shown in this wiki page: https://en.wikipedia.org/wiki/Meteorite_classification

- Stony meteorites
- Stony-iron meteorites
- Iron meteorites

Recclass

As we can see here, most of our meteorites are of the "Stony" kind



Mass

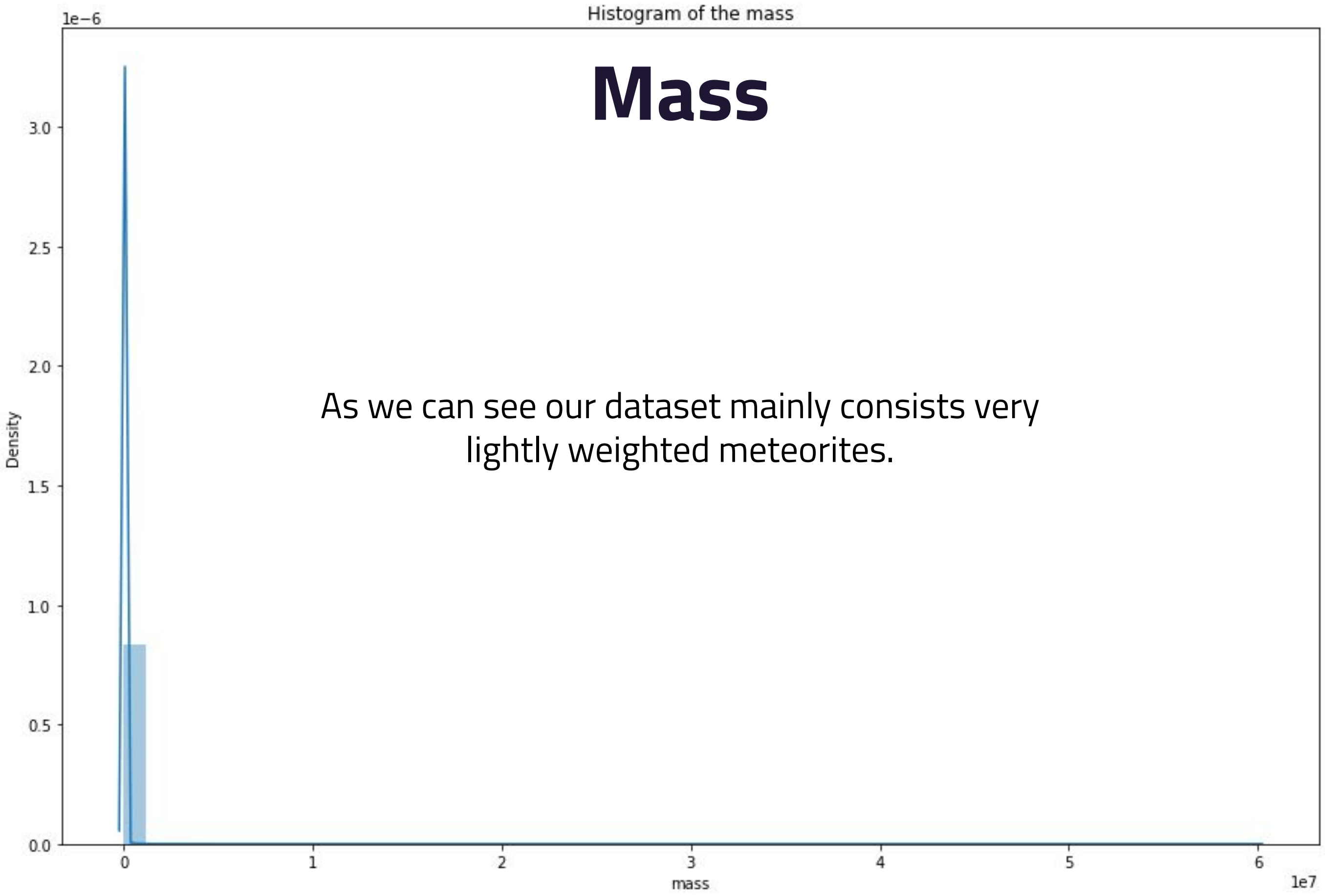
This column relates to the mass of the given meteorite. This mass is in grams.

- Lowest value is at 0 grams, max value at 60,000,000 grams (which is 60 tons)
- Mean is at 15,600 grams but the Median is at only 29.09 grams

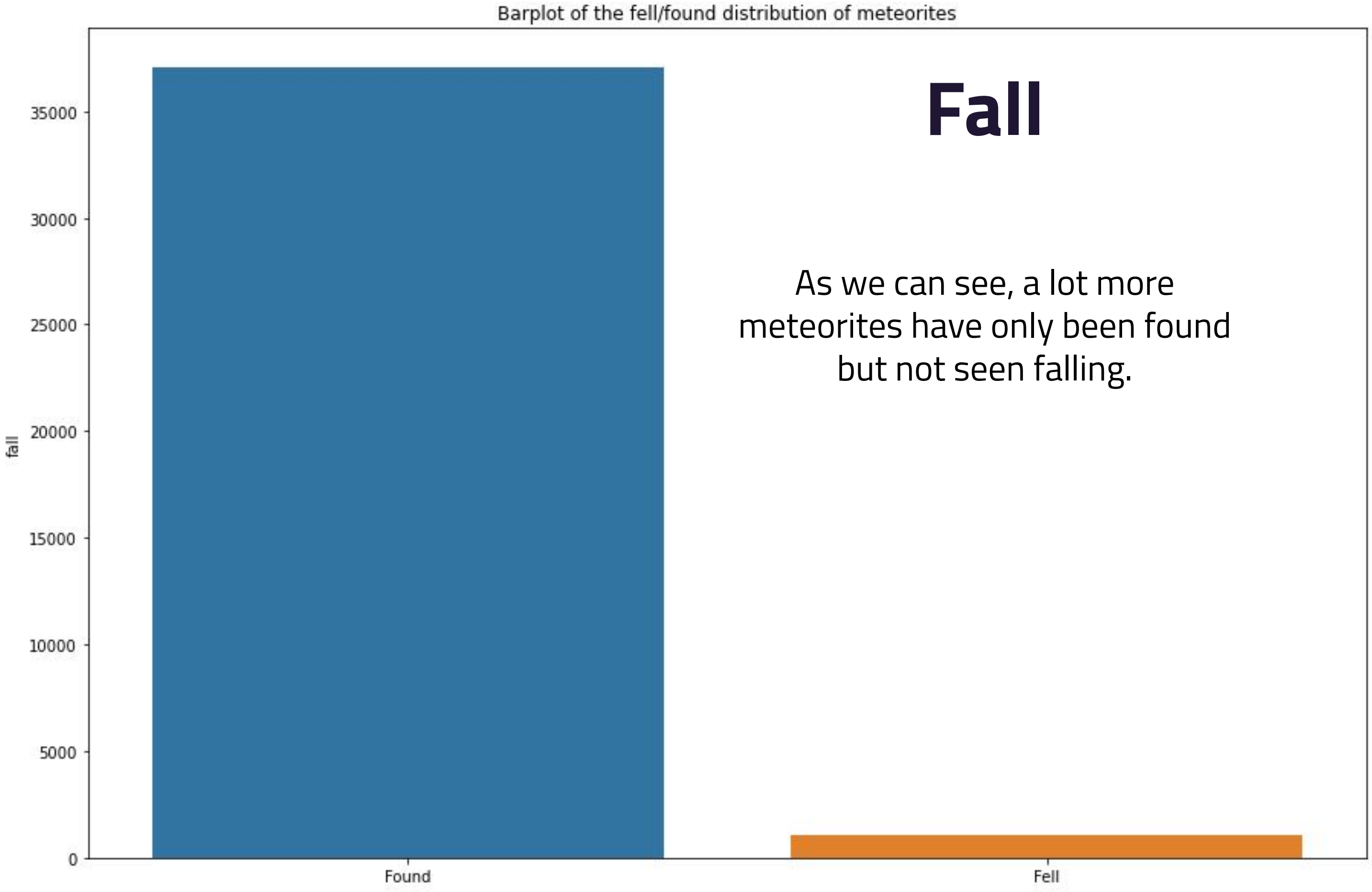
```
ds_mass.describe()

count    3.811500e+04
mean     1.560071e+04
std      6.286817e+05
min      0.000000e+00
25%      6.630000e+00
50%      2.909000e+01
75%      1.872900e+02
max      6.000000e+07
Name: mass, dtype: float64
```

Mass



As we can see our dataset mainly consists very lightly weighted meteorites.

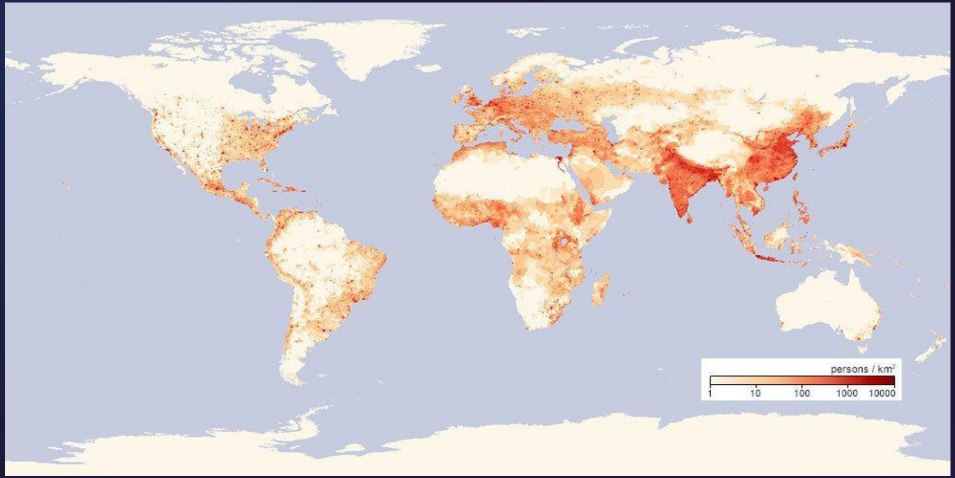


Fall

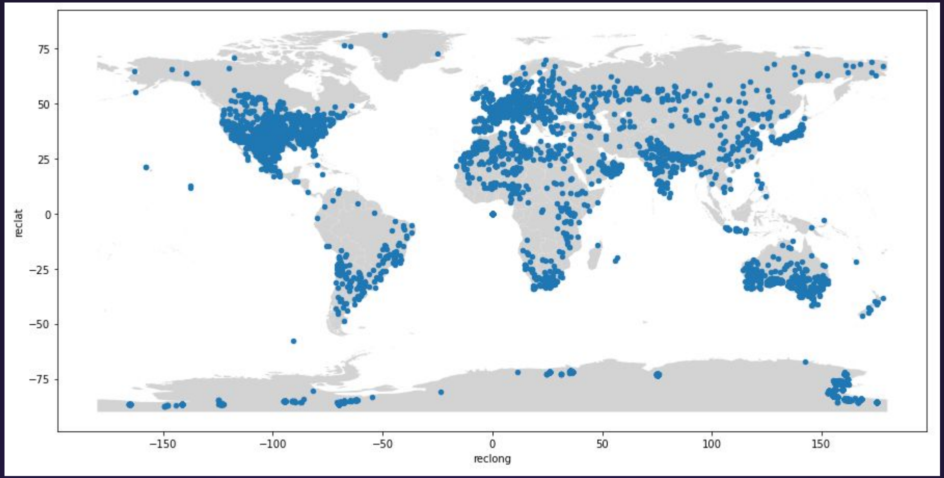
As we can see, a lot more meteorites have only been found but not seen falling.

Correlation between meteorites distribution and world population density

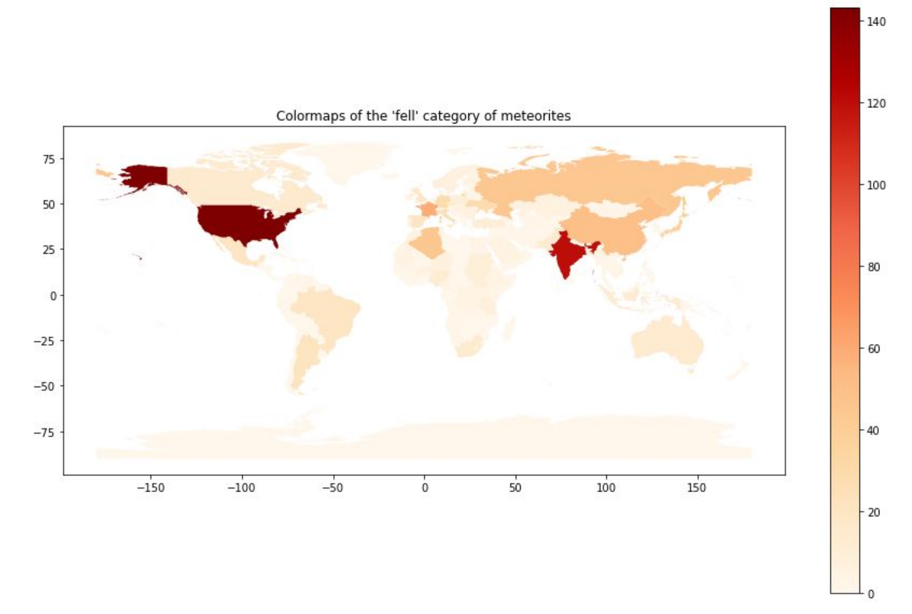
Heatmap by world population density



Map with all of the meteorite landings in our dataset



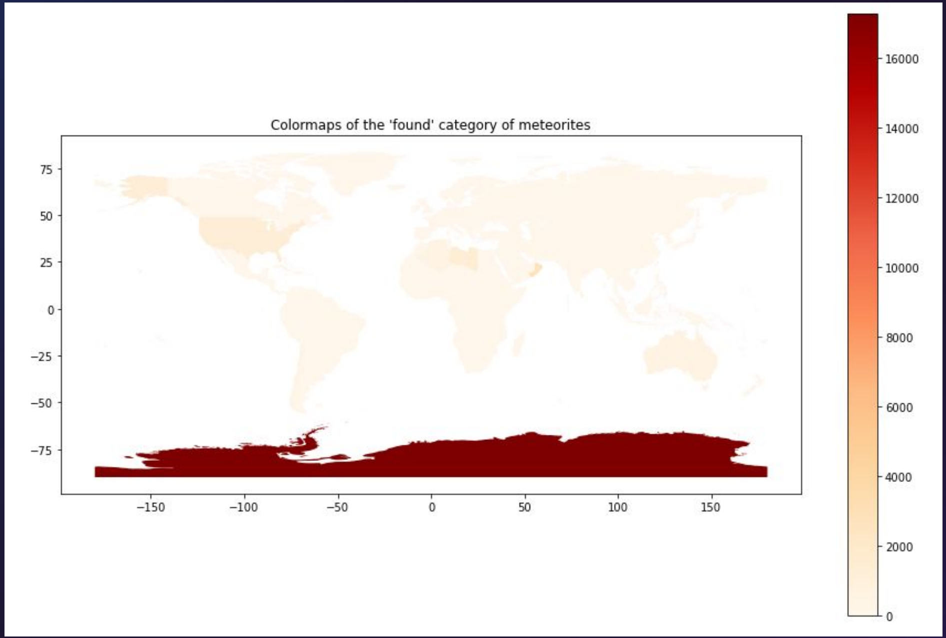
Heatmap of “fell” meteorites in different countries



Fell meteorites are meteorites that being seen fallen.

In the following heatmap we can see that there is correlation between the amount of “fell” meteorites to high populated countries (like India, China and the United States)

Are Meteorites more likely to be found in specific areas in the world?



The following heatmap describe the amount of “found” meteorites by country.

As we can see, most of the meteorites were found in Antarctica.

04 Machine Learning

Can we predict the year of landing based on meteorites coordinates?

To answer this question we used KNeighborsClassifier from ‘sklearn’ library and trained a model that predict the meteorite year of landing based on it coordinates.

We chose to use KNN in our scenario because of the fact that meteorites are usually landing in groups in a small distance from each other.

ML - Results

We were able to predict the year of landing in accuracy of 0.57

```
The accuracy is:  0.5714285714285714
Some examples of the guessed years, and the correct ones:
```

	pred	y_test
43051	1979-01-01T00:00:00.000	1986-01-01T00:00:00.000
17036	2007-01-01T00:00:00.000	2009-01-01T00:00:00.000
17066	2009-01-01T00:00:00.000	2009-01-01T00:00:00.000
8111	2000-01-01T00:00:00.000	2000-01-01T00:00:00.000
2104	1984-01-01T00:00:00.000	1985-01-01T00:00:00.000
...	...	...
4789	1988-01-01T00:00:00.000	1988-01-01T00:00:00.000
33401	1999-01-01T00:00:00.000	1994-01-01T00:00:00.000
2212	1985-01-01T00:00:00.000	1985-01-01T00:00:00.000
34680	1999-01-01T00:00:00.000	1999-01-01T00:00:00.000
37116	2002-01-01T00:00:00.000	2002-01-01T00:00:00.000

3171 rows × 2 columns

05 Conclusions

Conclusions

As we have seen, we were able to find out some interesting facts about recorded meteorites, like:

- Most of the meteorites are stony
- The majority of the meteorites are very lightweight (few grams)
- Most of the meteorites were found in Antarctica (mostly because of the unique climate)
- Most of the observed fell meteorites are in populated areas in the world
- Meteorites landing year can be predicted by coordinates (based on data of other meteorites that landed nearby)


THE END