Name: Naud Sluis
Student number: 14607883

Dataset:
https://archive.ics.uci.edu/ml/datasets/Dry+Bean+Dataset

Records: 13.611
Variables: 17
Dataset organization: Faculty of Technology, Selcuk University, Turkey

Description:
The dry-bean dataset was used in an experiment for the purpose of classifying dry beans using computer vision and machine learning techniques. The dataset contains sizes and shapes of seven types of beans types of beans. There are a total of twelve and four shape factors. The dataset contains fourteen variables

The variables in this dataset are:

- Area (Discrete, Ratio)
- Perimeter (Discrete, Ratio)
- Major axis length (Discrete, Ratio)
- Minor axis length (Discrete, Ratio)
- Aspect ratio (Discrete, Ratio)
- Eccentricity(Discrete, Ratio)
- Convex area(Discrete, Ratio)
- Equivalent diameter(Discrete, Ratio)
- Extent(Discrete, Ratio)
- Solidity(Discrete, Ratio)
- Roundness(Discrete, Ratio)
- Compactness(Discrete, Ratio)
- ShapeFactor1-4(Discrete, Categorical)
- Class(Discrete, Categorical)

Table 1. Descriptive statistics for each variable

Variable	Mean	Median	Standard Deviation	IQR	Mode
Area	53048.285	44652.0	29324.096	25004.0	N/A
Major axis	320.142	296.883	85.694	123.191	N/A
ShapeFactor3	0.644	0.642	0.099	0.115	N/A
ShapeFactor4	0.995	0.996	0.004	0.004	N/A
Variable	Value	Frequency	Proportion	Percentage	Mode
Class	Dermason	3546	3546/13.611	26.05%	Dermason
	Sira	2636	2636/13.611	19.37%	
	Seker	2027	2027/13.611	14.89%	
	Horoz	1928	1928/13.611	14.17%	
	Cali	1630	1630/13.611	11.98%	
	Barbunya	1322	1322/13.611	9.71%	
	Bombay	522	522/13.611	3.84%	

Research question:
Which shape factor has the most influence on bean-size?

Name: Naud Sluis
Student number: 14607883

Dataset:
https://zenodo.org/record/4770937

Records: 1000
Variables: 28/165
Organization: Structured
Dataset organization: European Climate Assessment & Dataset

Description:
This dataset consist of weather data from city's in Europe from 2001 to 2010. The dataset was originally created for machine learning purposes by Klein Tank, A.M.G et all. The weather data consists of multiple weather measurements like humidity, cloud coverage and temperature respectively.

The variables in this dataset are:

- Date (Continuous, Interval)
- Month (Discrete, Interval)
- Basel/Budapest/De_bilt-_cloud_cover(Discrete, Interval)
- Basel/Budapest/De_bilt-_humidity(Continuous, Ratio)
- Basel/Budapest/De_bilt-_pressure(Continuous, Ratio)
- Basel/Budapest/De_bilt-_global_radiation(Continuous, Ratio)
- Basel/Budapest/De_bilt-_precipitation(Continuous, Ratio)
- Basel/Budapest/De_bilt-_sunshine(Continuous, Ratio)
- Basel/Budapest/De_bilt-_temp_mean(Continuous, Interval)
- Basel/De_bilt-_temp_min(Continuous, Interval)
- Basel/Budapest/De_bilt-_temp_max(Continuous, Interval)

Table 1. Descriptive statistics for each variable

Variable	Mean	Median	Standard Deviation	IQR
Basel_Cloud_cover	5.439	6.0	2.369	3.0
Basel_min_temp	7.474	8.1	6.121	8.9
Basel_max_temp	16.052	16.6	8.11	12.625
Budapest_Cloud_cover	4.664	5.0	2.404	4.0
Budapest_mean_temp	12.804	13.95	8.62	13.

				925
Budapest_max_temp	12.272	18.5	9.816	15.9
De_Bilt_Cloud_cover	5.663	6.0	1.908	2.0
De_Blit_min_temp	6.867	7.4	5.406	8.0
De_Blits_max_temp	15.044	15.3	6.922	10.625

Research question:
What is the association between cloud-coverage and temperature?

Name: Naud Sluis
Student number: 14607883

Dataset:
https://statswales.gov.wales/Catalogue/Health-and-Social-Care/NHS-Hospital-Activity/nhs-activity-and-capacity-during-the-coronavirus-pandemic/admissions-by-date-patientype

Description search process:
I thought it would be interesting to look at hospital submissions in relation to for example covid-cases. I searched on the google dataset website for hospital submission per date, and this dataset was the result of that.

Chosen option:
This dataset contains information about (non)covid-19 hospital admissions in Wales. I will be doing option 1.

Table 1. Correlation overview

Variable from COVID-19 dataset	Variable from the Wales hosp_admissions dataset	Pearson's r
weekly_hosp_admissions	Non-COVID-19 admissions	-0.104
icu_patiets	COVID-19 admissions (suspected and confirmed)	0.106
hosp_patients	COVID-19 admissions (suspected and confirmed)	-0.192

Analysis description:
I thought about what was said on the news: 'The icu's are full'. I wanted to test this. I looked at hospital admissions to non-covid addmisions. They are indeed a bit lower, because people where more inside, so less accidents. I also looked at the icu patients to covid admissions, and there was little to no correlation. That seems different then what the news told.