

# Making Scraped Data Usable

---



**Ian Ozsvald**

INTERIM CHIEF DATA SCIENTIST

@ianozsvald ianozsvald.com



# Overview



**Export a CSV file**

**Create a Pandas DataFrame**

**Review our data**

**Investigate relationships graphically**

**Report to tell a visual story**



## Exploratory Data Analysis (EDA)

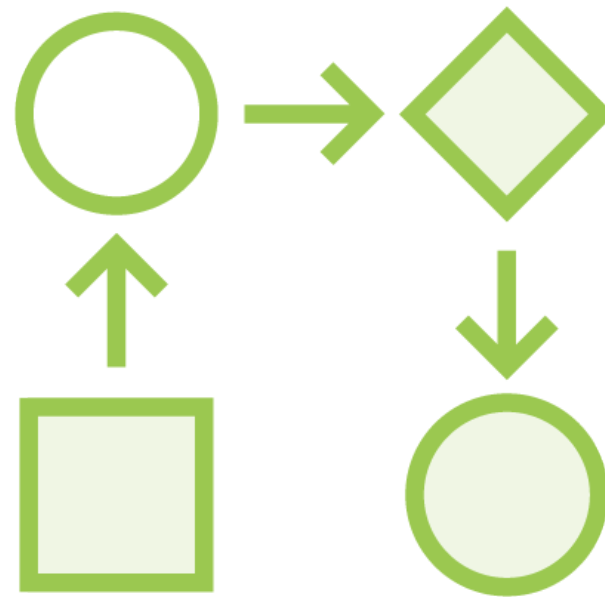
Missing data (NaNs)?

Infrequent items?

Stability over time?

Mis-spellings?

Visualise your  
columns



Weight and MPG?

Cylinders and  
Displacement?

MPG improvements  
over time?

Other interesting  
relationships?

Discover a story



Know your goals  
Check correlations  
Visualise  
Explore  
Keep it playful



# Summary



Scraped data using HTML elements and text processing

Reviewed, cleaned and discovered problems in our data

Explored relationships

Reported on our findings using a variety of charts

