



Internship Presentation

Daniel Odeleye

Project Definition

Purpose:

- **Robustness:** Analysing integrity and accuracy of fundamental datasets, to identify issues that may exist in the data.
- **Jacobsen:** Analysing replacement feed data for Jacobsen – matching old to new metadata

Solutions:

- **Robustness:** Created a module that takes as input the name of a fundamental dataset and performs test for its robustness.
- **Jacobsen:** Created a script which uses a mapper file to match two Jacobsen sources with same time-series but represented by different metadata.

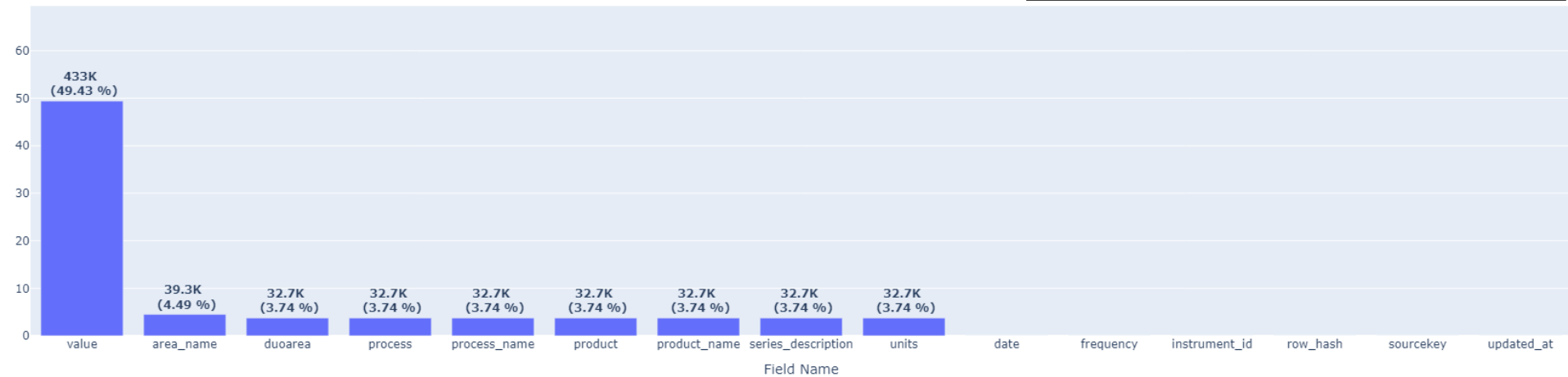
Challenges:

- **Robustness:** Extracting large datasets, complexity of test cases, creating appropriate data visualisations using a new library, identifying frequency of each time series.
- **Jacobsen:** Handling Unicode characters in metadata, null values, duplicate mappings.

Visualisations 1 – “eia-weekly-psr”

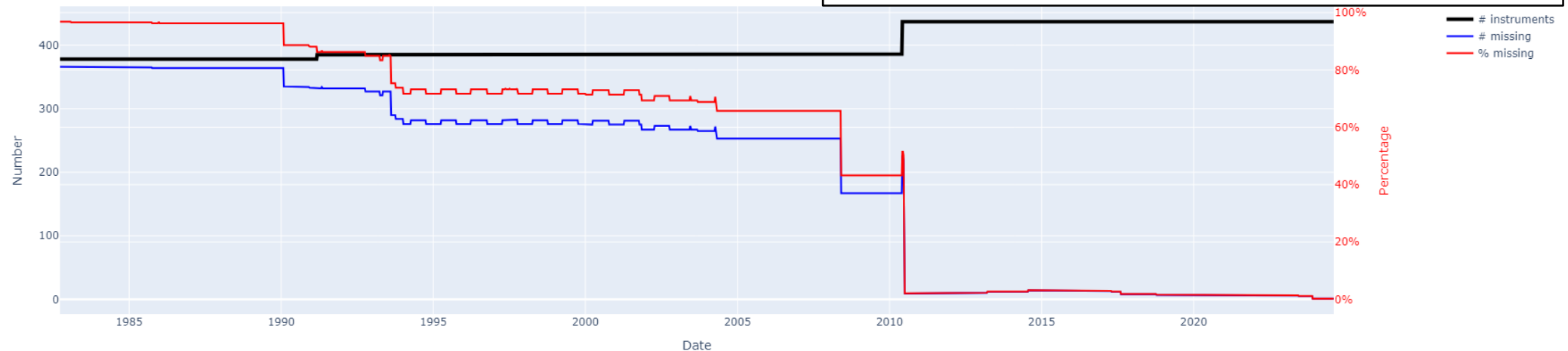
Percentage of missing values per column in 'eia-weekly-psr', 4-Week

Number of nulls per column



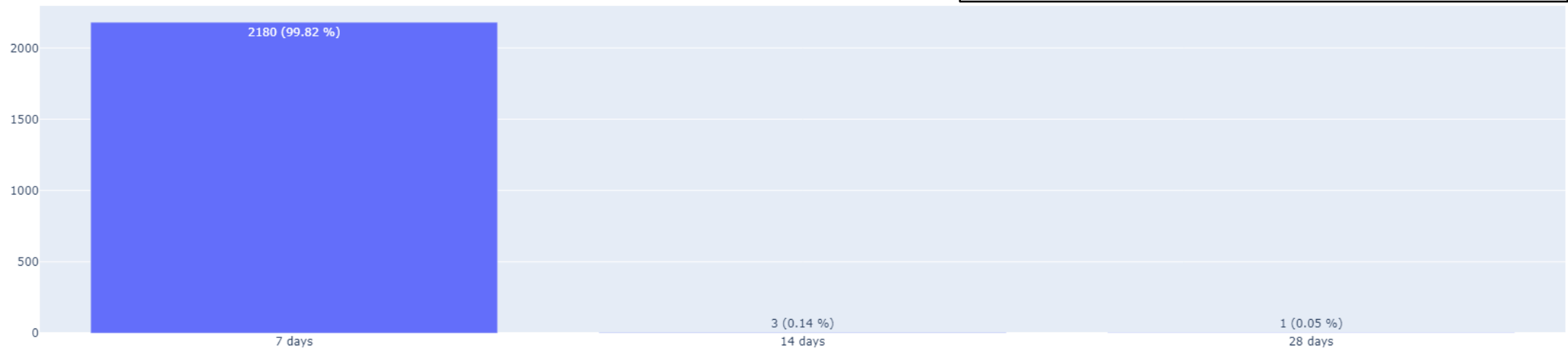
Number of missing values (under the 'value' field) over time - 'eia-weekly-psr', 4-Week

Number of instruments & null values per date

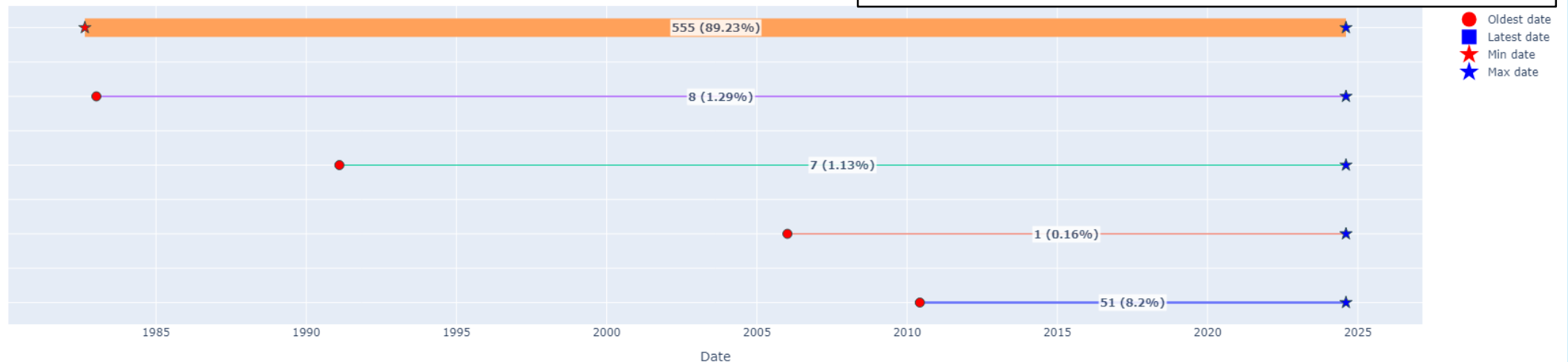


Visualisations 2 – “eia-weekly-psr”

Distribution of interval lengths between dates - 'eia-weekly-psr', Weekly



Date ranges per instrument - 'eia-weekly-psr', Weekly



Outputs – “Jacobsen” Matchups

Instrument matching between vendors

symbol	pricetype	hash_for_prod	full_metadata	sym_price_in_jacobsen_api	sym_price_in_fastmarkets_api	sym_in_fastmarkets_api
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	TRUE	TRUE	TRUE
AG-BM-0001	PriceHigh	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceHigh,nan,nan	Animal Fats and Oils	TRUE	TRUE	TRUE
AG-BM-0001	PriceLow	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceLow,nan,nan	Animal Fats and Oils	TRUE	TRUE	TRUE
AG-BM-0002	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,439,PriceAverage,nan,nan	Animal Fats and Oils	TRUE	TRUE	TRUE
AG-BM-0002	PriceHigh	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,439,PriceHigh,nan,nan	Animal Fats and Oils	TRUE	TRUE	TRUE
AG-BM-0002	PriceLow	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,439,PriceLow,nan,nan	Animal Fats and Oils	TRUE	TRUE	TRUE
AG-BM-0003	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,178,457,PriceAverage,nan,nan	Animal Fats and Oils	TRUE	TRUE	TRUE
AG-BM-0003	PriceHigh	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,178,457,PriceHigh,nan,nan	Animal Fats and Oils	TRUE	TRUE	TRUE
AG-BM-0003	PriceLow	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,178,457,PriceLow,nan,nan	Animal Fats and Oils	TRUE	TRUE	TRUE
AG-CCN-0001	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,99,93,PriceAverage,nan,nan	Vegetable Oils,344a15	TRUE	TRUE	TRUE

Date-value pair comparison by instrument

symbol	pricetype	hash_for_prod	full_metadata	date	value_prod	value_dev	match
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	22/03/2002	985	985	TRUE
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	04/04/2002	985	985	TRUE
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	11/04/2002	985	985	TRUE
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	18/04/2002	985	985	TRUE
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	25/04/2002	985	985	TRUE
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	02/05/2002	965	965	TRUE
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	09/05/2002	945	945	TRUE
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	16/05/2002	945	945	TRUE
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	23/05/2002	940	940	TRUE
AG-BM-0001	PriceAverage	Animal Fats and Oils - International,eee201fc-e77f-4ec6-8bce-67edf206b621,166,450,PriceAverage,nan,nan	Animal Fats and Oils	30/05/2002	870	870	TRUE

Outputs – “Jacobsen” Matchups

symbol	pricetype	hash_for_prod	full_metadata	date	value_prod	value_dev	match
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	02/05/2022	80.09	6	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	03/05/2022	80.28	6	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	04/05/2022	82.43	6	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	05/05/2022	81.85	5.5	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	06/05/2022	80.9	5.5	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	09/05/2022	79.74	5.5	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	10/05/2022	81.04	5.5	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	11/05/2022	83.45	5.5	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	12/05/2022	82.52	4	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	13/05/2022	83.79	4	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	16/05/2022	82.99	4	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	17/05/2022	83.49	4	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	18/05/2022	80.55	5	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	19/05/2022	79.53	5	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	20/05/2022	80.93	5	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	23/05/2022	80.47	4	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	24/05/2022	80.12	4	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	25/05/2022	78.92	4	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	26/05/2022	80.52	4	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	27/05/2022	79.57	4	FALSE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	30/05/2022	79.57	79.57	TRUE
AG-SYB-0046	PriceAverage	Vegetable Oils,344a1556-d0e2-41ab-9614-3000a669dd4f,96,20048,PriceAverage,nan,nan	Vegetable Oils,344a15	31/05/2022	77.92	4	FALSE

Project Implementation

Use **pandas** DataFrames for data manipulation and checking against cases

Save findings to **CSV** files from pandas DataFrames

