

Racing Towards Insights

*Analysing the Sale Sizzler 5k Series with
VSCode Data Wrangler And Power BI*

Damien Jones (he/him)
Data Engineer, SecureDev



MrDamienJones



MrDamienJones



amazonwebshark.com



damien@amazonwebshark.com



@amazonwebshark

Damien Jones

♂ He/Him 🌐 Manchester UK 🦈 Fin Fan

Data Engineer @ SecureDev

AWS & Azure Certified

Creator @ amazonwebshark.com

Runner; Keen Gardener; Dog Dad



Agenda

Sale Sizzlers Introduction

VS Code Data Wrangler Introduction

Data Wrangler & Power BI Demos

Summary & Questions



Sale Harriers
Manchester

Sale Sizzlers

5k road race series operated by Sale Harriers
Manchester

First events in 2002

Four events each Summer

Since 2016 results published by Nifty as CSV



Sale Sizzlers Data

Runner Data

Number	First name	Last name	Club
490	Thomas	Charles	Chorlton Runners
125	Ryan	Moore	Vale Royal AC
498	Martin	Clark	Bury Athletic Club
487	Harry	Johnson	Halifax Harriers & AC
183	Paul	Peters	Sale Harriers Manchester

Event Data

Finish time	Chip time
00:15:11	00:15:11
00:15:26	00:15:26
00:15:29	00:15:28
00:15:34	00:15:33
00:15:59	00:15:58
00:16:05	00:16:04

Sale Sizzlers: Rank Data

Overall Position

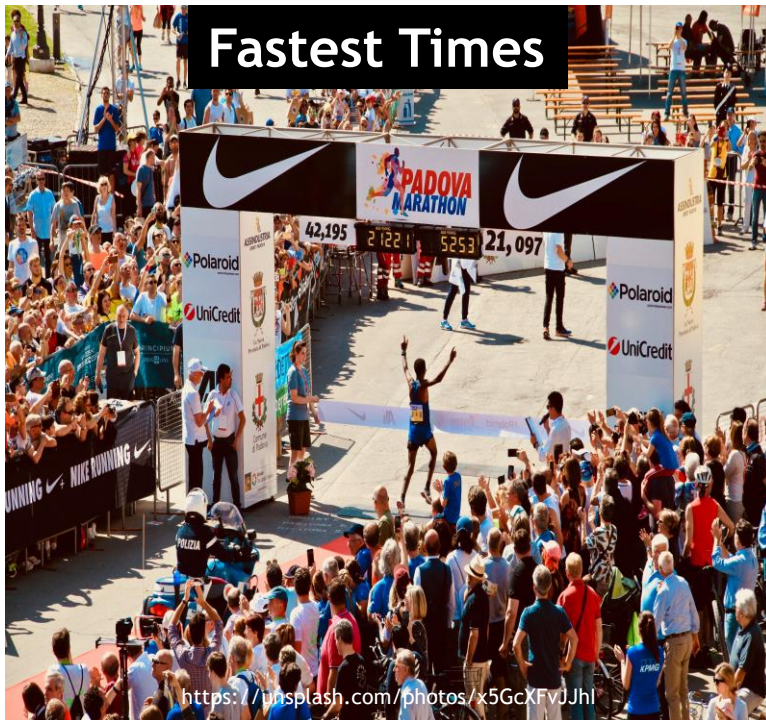
Gender Position

Age Category Position

Position	Finish time	Gender	Gender position	Category	Category position
1	00:15:11	Male	1	Senior Male (20-39)	1
2	00:15:26	Male	2	Senior Male (20-39)	2
3	00:15:29	Male	3	Senior Male (20-39)	3
4	00:15:34	Male	4	Senior Male (20-39)	4
5	00:15:59	Male	5	Senior Male (20-39)	5
6	00:16:05	Male	6	Under 20	1
7	00:16:05	Male	7	Under 20	2

Insights

Fastest Times



Gender & Age



Running Clubs



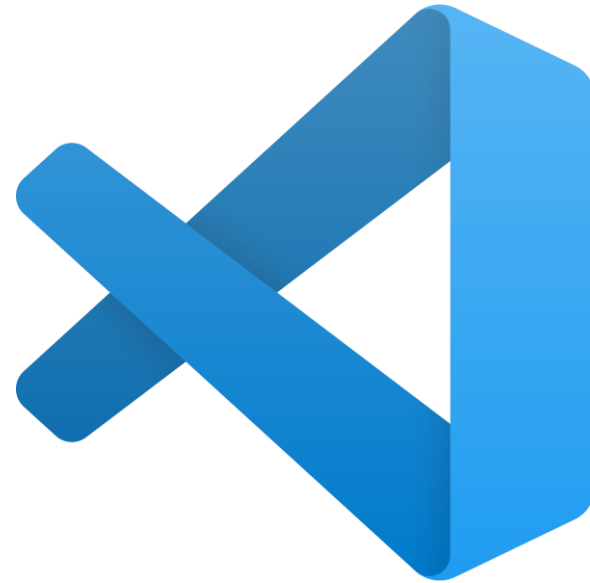
Visual Studio Code

Read Code

Write Code

Deploy Code

Free!



VS Code Data Wrangler

VS Code Extension by Microsoft

In active development

Designed for data cleansing

Also free!



Data Wrangler Benefits

No-code interface

Data updates in real time

Source data unchanged

Similarities to Excel and Power Query

Data Wrangler Benefits

One single transformation process

Data is ready when it hits Power BI

Reduce Power BI computational load

Transform data closer to source

Roche's Maxim of Data Transformation

Data should be transformed...

As far **upstream** as possible

As far **downstream** as necessary



OPERATIONS

Find and replace (4)

- Drop duplicate rows
- Drop missing values
- Fill missing values
- Find and replace

> Format (7)

> Formulas (4)

> Numeric (4)

> Schema (5)

> Sort and filter (2)

Custom operation

DATA SUMMARY

Data shape 548 rows x 14 columns

Columns 14

> Rows 548

> Missing Values (by column) 1,123

CLEANING STEPS

1 Loaded variable 'df' from URI: /home/damien/d...

2 New operation

[Export to Notebook](#) [Export as CSV](#) [Copy all code](#)[Report an issue](#)

index

Position

Finish time

Number

First name

Missing: 132 (24%)
Unique: 409 (75%)Missing: 0 (0%)
Unique: 333 (61%)Missing: 131 (24%)
Unique: 417 (76%)Missing: 0 (0%)
Unique: 263 (48%)Not started:
00:17:54:
00:22:00:
Other:263
Unique values

0	1	00:15:11	490	Thomas
1	2	00:15:26	125	Ryan
2	3	00:15:29	498	Martin
3	4	00:15:34	487	Harry
4	5	00:15:59	183	Paul
5	6	00:16:05	197	Finlay
6	7	00:16:05	679	George
7	8	00:16:08	455	Jacob
8	9	00:16:10	483	Tommy
9	10	00:16:13	646	Harry
10	11	00:16:20	476	Paul
11	12	00:16:26	629	Sonny
12	13	00:16:28	456	Tom
13	14	00:16:29	138	Martin

PROBLEMS

TERMINAL

PORTS

7

DATA WRANGLER

COMMENTS

DEBUG CONSOLE

OUTPUT

2 New operation



1 Choose an operation to get started

Or, start typing code to see a live preview of the transformation on your data (e.g. `df =`


Choose an operation to generate preview

Apply

Discard

# index	# Position	Finish time	# Number	First name
Missing: 132 (24%)	Missing: 132 (24%)	Missing: 0 (0%)	Missing: 131 (24%)	Missing: 0 (0%)
Unique: 409 (75%)	Unique: 409 (75%)	Unique: 333 (61%)	Unique: 417 (76%)	Unique: 263 (48%)
		Not started: 00:17:54: 00:22:00: Other:		263 Unique values
Min 1 Max 416				
		Min 101 Max 737		

A screenshot of a software interface showing a menu titled 'OPERATIONS'. At the top of the menu is a search bar with a magnifying glass icon and the text 'Search for operations...'. Below the search bar, the menu is expanded to show several options. The first option is 'Find and replace (4)', which is further expanded to show four sub-options: 'Drop duplicate rows', 'Drop missing values', 'Fill missing values', and 'Find and replace'. Below these are four more main options, each preceded by a right-pointing chevron: 'Format (7)', 'Formulas (4)', 'Numeric (4)', and 'Schema (5)'. At the bottom of the menu is the option 'Sort and filter (2)'. The entire menu is set against a light gray background.

 Search for operations...

Find and replace (4)

Drop duplicate rows

Drop missing values

Fill missing values

Find and replace

> Format (7)

Formulas (4)

> Numeric (4)

> Schema (5)

> Sort and filter (2)

Custom operation

The figure consists of two side-by-side bar charts. Both charts have 'Size Class' on the x-axis and 'Number of Species' on the y-axis. The y-axis ranges from 0 to 100. The left chart shows a high number of species in the smallest size class (around 100) and a sharp decline in larger classes. The right chart shows a more even distribution across size classes, with a peak in the middle (around 50).

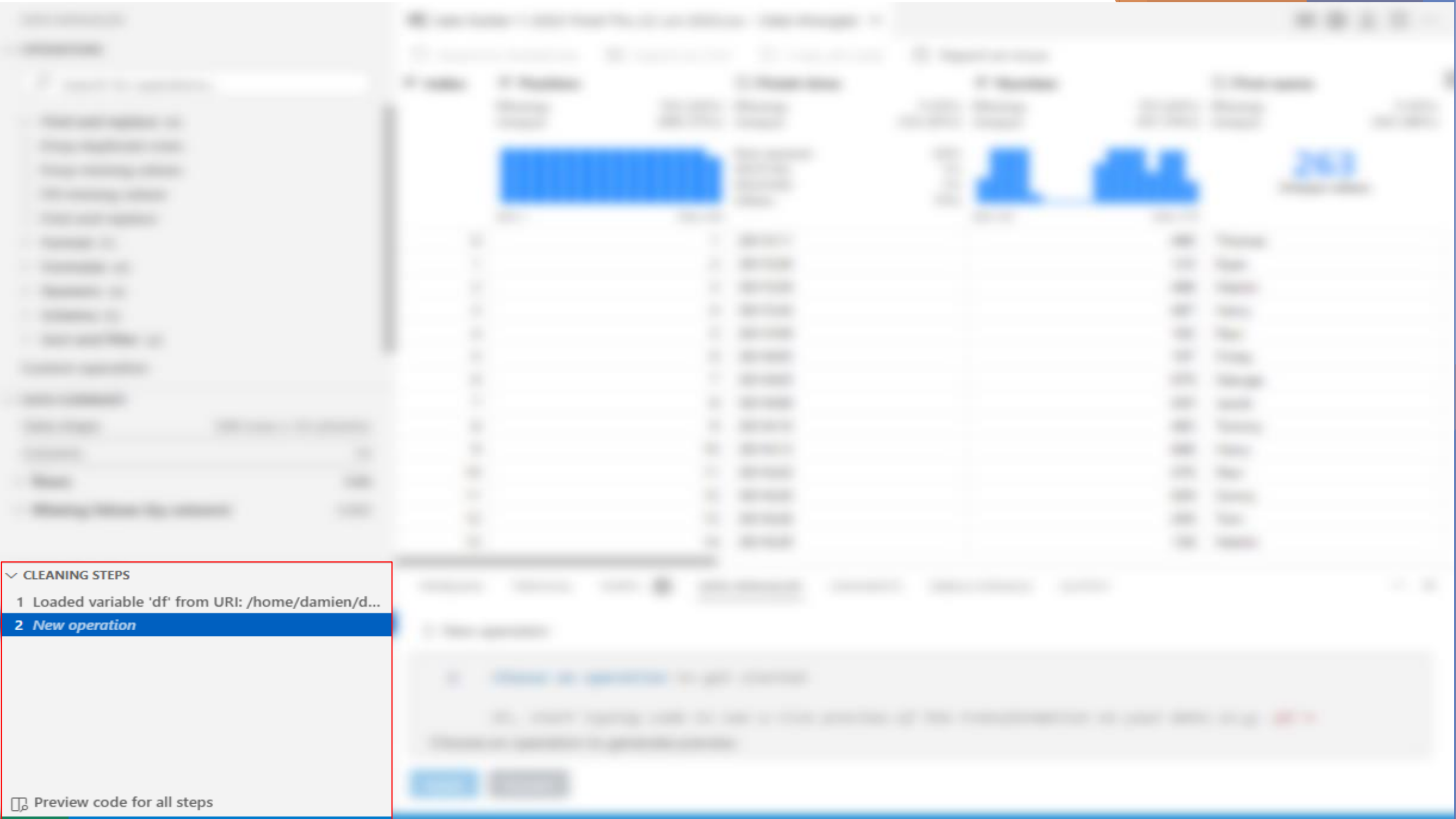
DATA SUMMARY	
Data shape	548 rows x 14 columns
Columns	14
> Rows	548
> Missing Values (by column)	1,123

▼ CLEANING STEPS

1 Loaded variable 'df' from URI: /home/damien/d...

2 *New operation*

Preview code for all steps





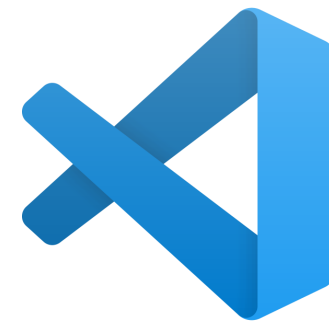
2 New operation

1 Choose an operation to get started

Or, start typing code to see a live preview of the transformation on your data (e.g. `df =`
Choose an operation to generate preview

Apply

Discard



Demos - VS Code (Sizzler 1)



Demos - Power BI

Summary

Sale Sizzlers

What they are

What data they generate

What insights were wanted

VS Code Data Wrangler

What are VS Code and Data Wrangler

Data Wrangler benefits

Data Wrangler features

Demo

Power BI

Built a model

Examined visuals

Generated insights

FINISH
Congratulations!

Me



Thanks!



MrDamienJones



MrDamienJones



amazonwebshark.com



damien@amazonwebshark.com



@amazonwebshark

<https://www.flaticon.com/>