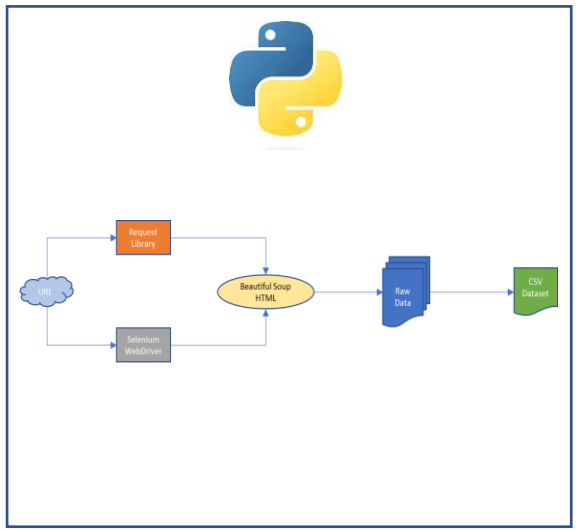
jupyter

Azure

Power BI

Web scraping program is written in Python 3 using Jupyter Notebook



Data scraped from:

- https://www.wine-searcher.com/critics-8-cellartracker?page=1
- https://www.winemag.com/ratings/?s=&drink_type=wine&sort_by=pub_date_web&sort_dir=desc

Azure Blob Storage

Dataset is stored in Azure Blob Storage as CSV

Azure Synapse Analytics

Dataset is ingested in Azure Synapse Analytics, transformed and sinked into an SQL Pool

Copy Data Transform Data

Data is pulled from the SQL Pool in Azure Synapse and integrated with Power BI

Different dashboards are built in Power BI for analysis

Overview on the raw data scraped:

	Product_Name	Appellation	Rating	Price
0	Charles Heidsieck NV Champagne Charlie Brut (...	Champagne	100	$700
1	Stag's Leap Wine Cellars 2019 Fay Estate Grown...	Napa	100	$155
2	Marqués de Riscal 2016 150 Aniversario Gran Re...	Northern Spain	99	$60
3	Salon 2012 Le Mesnil Blanc de Blancs Brut Char...	Champagne	99	$992
4	Philipponnat 2013 Clos des Goisses Extra Brut ...	Champagne	99	$320
...	...	...	...	...
11099	Bodegas Franco-Españolas 2016 Royal Reserva Te...	Northern Spain	92	$45
11100	Cedarville 2019 Estate Grenache (El Dorado)	Sierra Foothills	92	$36
11101	Doña Paula 2018 Alluvia Parcel Bush Vines Malb...	Mendoza Province	92	$100
11102	El Enemigo 2017 Gran Enemigo El Cepillo Single...	Mendoza Province	92	$100
11103	V. Sattui 2017 Prestige Cuvée Blanc de Noir Sp...	Mendocino County	92	$52

11104 rows × 4 columns

	Scoring	Scale
0	98–100	Extraordinary
1	94–97	Outstanding
2	90–93	Very good
3	86–89	Good
4	80–85	Average
5	70–79	Below average
6	50–69	Avoid

The above dataset is scraped from 1850 pages