

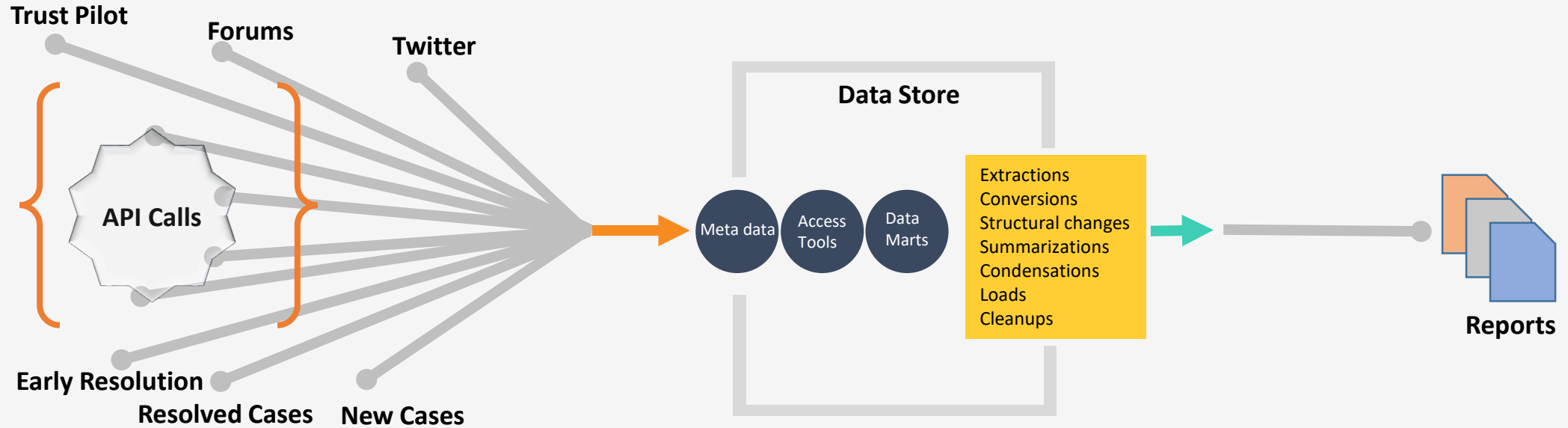
Financial Conduct Authority

Data Engineering and Analytics

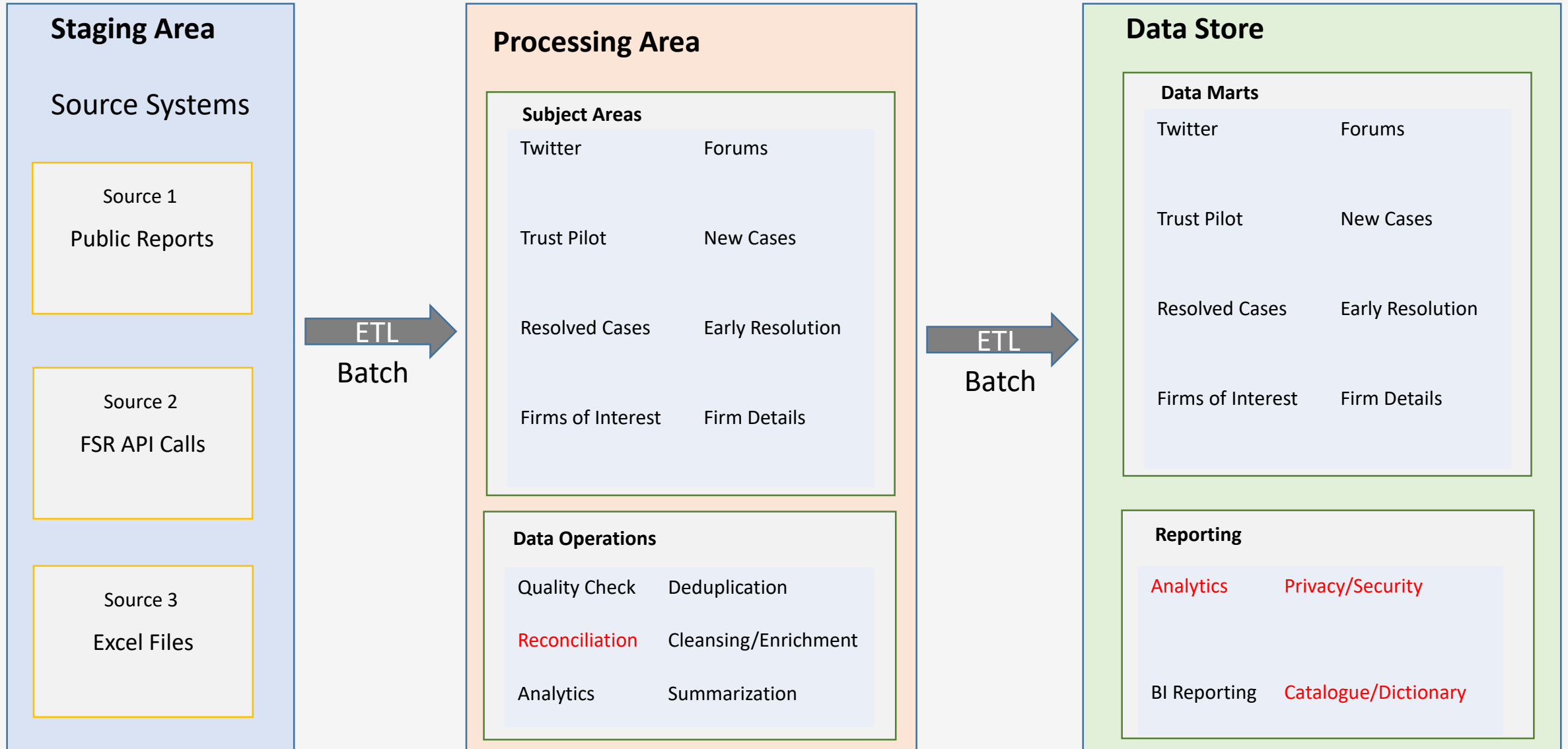
Interview Round II

Muzaffar Mohammed

Data Engineering Overview



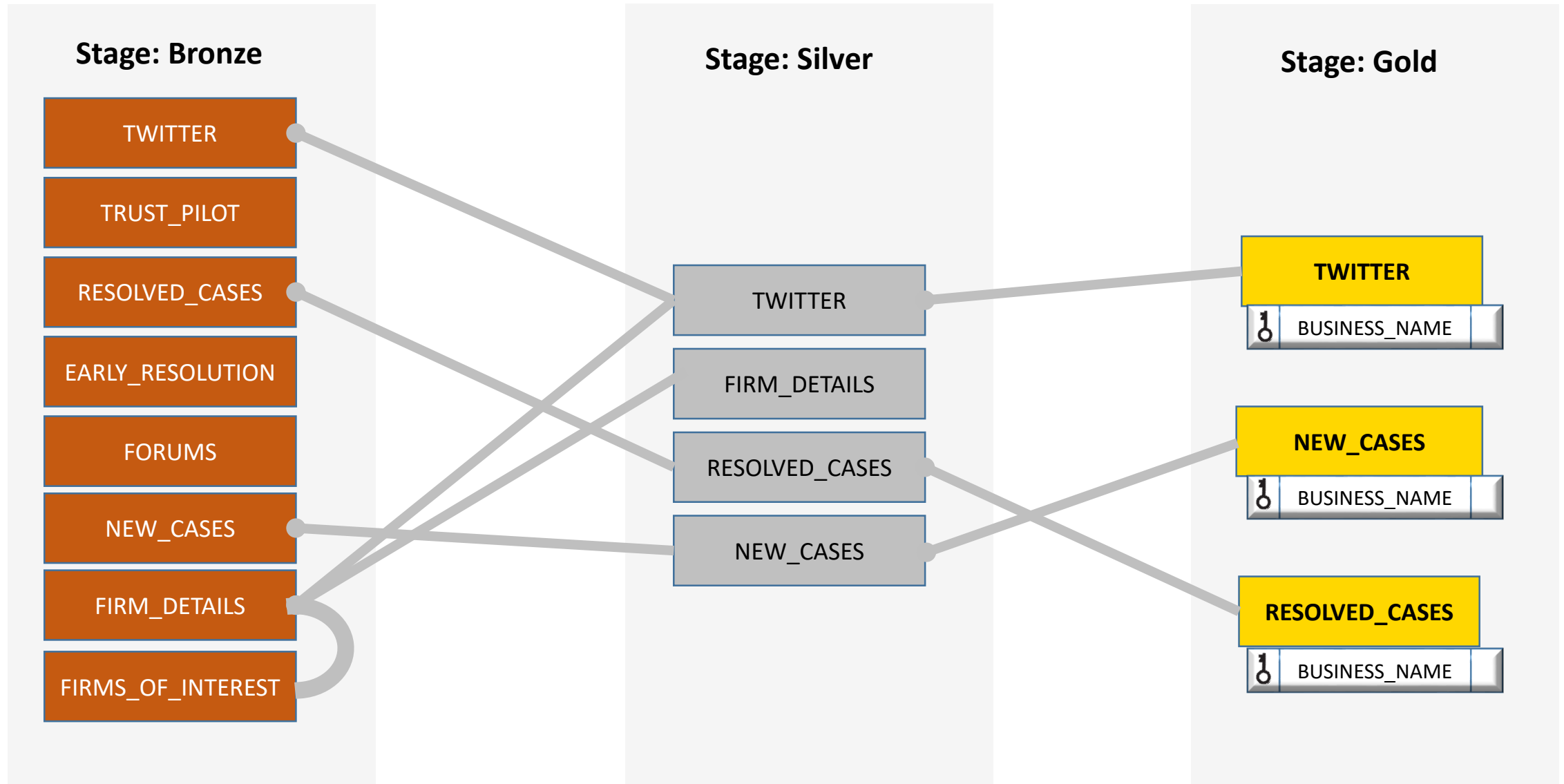
Data Flow – 3 Tier Architecture



Data Engineering – OEM Products



Data Store – Data Models



Data Profiling – Data Cleansing

Data Profiling: Findings

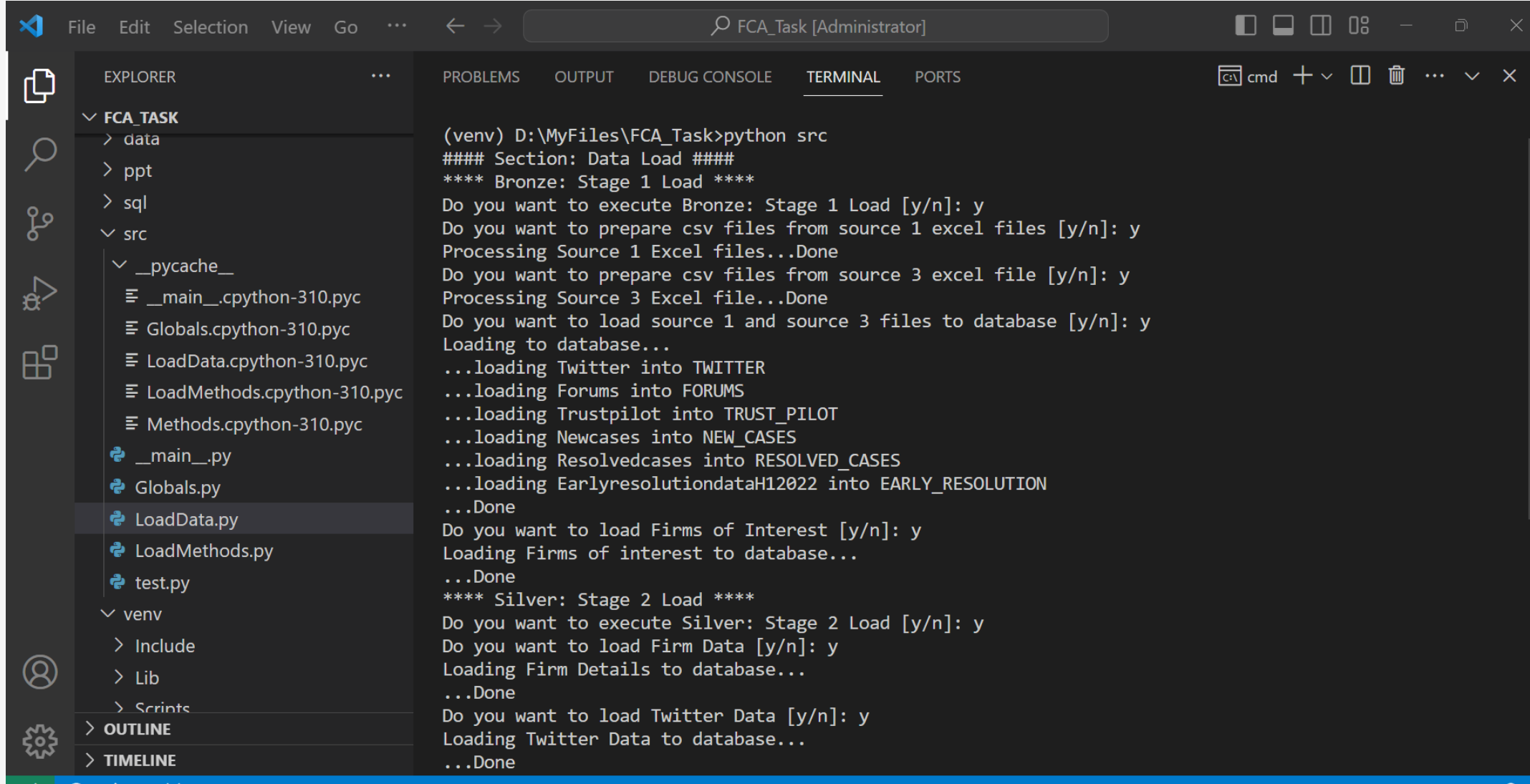
Firm Reference Number is missing

1. Financial Services Register API mandates Firm Reference Number in all API calls
2. Efforts to use Tw Despite attempts to join the three sources using the mentioned approach and generate data analytics, the outcome produced limited data, and the accuracy of this data is questionable tweet Tags as Firm Name via Common Search API resulted in challenges, including multiple Business Names or no output for some Tags
3. A more accurate and reliable approach is required for resolving this matter

Unable to link all 3 sources

1. Due to the aforementioned reason, we encountered challenges in successfully integrating all three sources
2. Despite attempts to join the three sources using the mentioned approach and generate data analytics, the outcome produced limited data, and the accuracy of this data is questionable

Data Loading – In action



The screenshot shows a Visual Studio Code interface with a terminal window open. The Explorer pane on the left shows a project named 'FCA_TASK' with a 'src' folder containing several Python files. The terminal window shows the execution of a Python script, with prompts for confirmation to execute different stages of data loading. The script is running in a virtual environment at 'D:\MyFiles\FCA_Task'.

```
(venv) D:\MyFiles\FCA_Task>python src
#### Section: Data Load ####
**** Bronze: Stage 1 Load ****
Do you want to execute Bronze: Stage 1 Load [y/n]: y
Do you want to prepare csv files from source 1 excel files [y/n]: y
Processing Source 1 Excel files...Done
Do you want to prepare csv files from source 3 excel file [y/n]: y
Processing Source 3 Excel file...Done
Do you want to load source 1 and source 3 files to database [y/n]: y
Loading to database...
...loading Twitter into TWITTER
...loading Forums into FORUMS
...loading Trustpilot into TRUST_PILOT
...loading Newcases into NEW_CASES
...loading Resolvedcases into RESOLVED_CASES
...loading EarlyresolutiondataH12022 into EARLY_RESOLUTION
...Done
Do you want to load Firms of Interest [y/n]: y
Loading Firms of interest to database...
...Done
**** Silver: Stage 2 Load ****
Do you want to execute Silver: Stage 2 Load [y/n]: y
Do you want to load Firm Data [y/n]: y
Loading Firm Details to database...
...Done
Do you want to load Twitter Data [y/n]: y
Loading Twitter Data to database...
...Done
```

Analytics – Sample Report



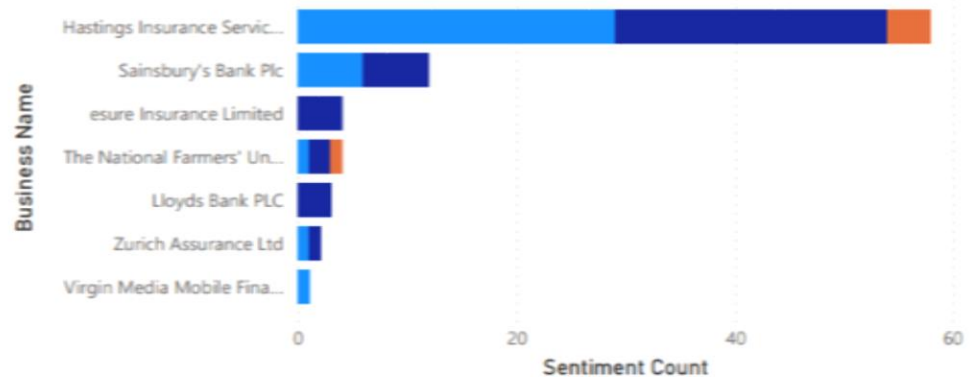
Half yearly Business Complaints - Jan - Jun 2022

Business Name	Total Cases	Resolved Cases
Zurich Insurance PLC	127.0	1
Yorkshire Building Society	67.0	2
Wise Payments Limited	174.0	1
West Bay Insurance Plc	193.0	1
Volkswagen Financial Services (UK) Limited	636.0	32
Virgin Media Mobile Finance Limited	59.0	1
Vanquis Bank Limited	874.0	33
Vanguard Asset Management, Ltd	70.0	2
Ulster Bank Ltd	30.0	1
U K Insurance Limited	928.0	45
TSB Bank plc	822.0	95
Think Money Limited	44.0	2
The Royal Bank of Scotland Plc	700.0	67
The Prudential Assurance Company Limited	224.0	2
The National Farmers' Union Mutual Insurance Society Limited	129.0	2
The Co-Operative Bank Plc	346.0	14
Tesco Personal Finance PLC	428.0	4
Starling Bank Limited	512.0	42
Standard Life Assurance Limited	58.0	1
St. James's Place Wealth Management Plc	89.0	3
Soteria Insurance Limited	49.0	1
Society of Lloyd's	163.0	6
Shop Direct Finance Company Limited	469.0	1
Shawbrook Bank Limited	672.0	6
Scottish Widows Limited	327.0	5
Scottish Equitable plc	94.0	4
Centender UK plc	2672.0	163

Business Name	Total Cases	Sentiment	Sentiment Count
esure Insurance Limited	312.0	neutral	4.00
Hastings Insurance Services Limited	247.0	negative	29.00
Hastings Insurance Services Limited	247.0	neutral	25.00
Hastings Insurance Services Limited	247.0	positive	4.00
Lloyds Bank PLC	2578.0	neutral	3.00
Sainsbury's Bank Plc	272.0	negative	6.00
Sainsbury's Bank Plc	272.0	neutral	6.00
The National Farmers' Union Mutual Insurance Society Limited	129.0	negative	1.00
The National Farmers' Union Mutual Insurance Society Limited	129.0	neutral	2.00
The National Farmers' Union Mutual Insurance Society Limited	129.0	positive	1.00
Virgin Media Mobile Finance Limited	59.0	negative	1.00
Zurich Assurance Ltd	108.0	negative	1.00
Zurich Assurance Ltd	108.0	neutral	1.00
Total			84.00

Sentiment Count by Business Name and Sentiment

Sentiment ● negative ● neutral ● positive



THANK YOU