Excel Project: Thought Process while making the project

Detailed Process Logbook

Excel Project: Investigating Differences Between IPOs Before and After 2021

Detailed Process Logbook and Analysis Report

---

1. Aim

To investigate the differences between IPOs launched **before and after 2021**, focusing on listing gains, issue size, and overall performance using Excel and Power Query.

---

2. Data Collection

- **Initial Source:** Kaggle dataset.
- **Problem Identified:** The dataset was **outdated** and lacked updated listing gains.
- **Solution Implemented:**
    - Learned **basic principles of IPO listings** and related terminology.
    - Used the following website for updated data:

🔗 https://www.capitalmarket.com/markets/IPOs/ipo-historic-table

    - Direct data fetching was **not possible**; used **Power Query (M Language)** to fetch the web data.

---

3. Data Availability

- **Total Companies:**
    - **BSE Listed:** 5,595
    - **NSE Listed:** 2,629
    - **Combined Total:** ~7,500 companies

(Source: List of Stocks on the Bombay Stock Exchange (BSE) in India)

---

4. Using Power Query for Web Data Extraction

- Initially, Power Query could only **pull one table** at a time from the webpage.

- To overcome this limitation, used the **function method** to retrieve multiple tables.
- Followed tutorial:

🔗 Scrape Data From Multiple Web Pages into Excel (with Power Query)

- The video helped extract **real-time data** corresponding to the collection date.
- **Problem Encountered:**
    - When writing the Power Query function, the same table appeared regardless of the variable input.
- **Troubleshooting:**
    - Used ChatGPT to debug the *M language* script.
    - Took approximately **5 hours** to get a working version.

---

5. *CMP Retrieval Issues*

- Attempted to find **Current Market Price (CMP)** for each company using:
- =XLOOKUP("*"&TRIM(B5)&"*",
- 'MARKET PRICE (2)'!B:B,
- 'MARKET PRICE (2)'!C:C,
- "Not found",
- 2)
-
- **Problem:**
    - VLOOKUP/XLOOKUP failed for many companies because company names did not match exactly (e.g., “UTI” and “Gland Pharma”).
- **Alternative Attempt:**

Used only the **first word of the company name** to match:

(LEFT(B3,FIND(" ",B3)-1),B3)

or

=IFERROR(LEFT(B3,FIND(" ",B3)-1),B3)

    - This helped retrieve some missing values, but mismatched companies (like “ICICI Bank” vs. “ICICI Lombard”) caused **incorrect CMP results**.

- **Conclusion:**

Using the first word method was **not reliable**.

Some anomalies remained in the data, especially for companies not listed in the market price table.

---

6. Handling Name Irregularities

- Tried removing unwanted suffixes like “Ltd” or “Lt” using:
- =TRIM(SUBSTITUTE(SUBSTITUTE(B2,"Ltd",""),"Lt",""))
-
- **Observation:** This formula didn’t clean names effectively for all cases.
- **Decision:**

Ignored minor mismatches and continued analysis using available listing gains, focusing on overall trends.

---

7. Pivot Table Creation

- Ignored CMP temporarily and focused on **listing gains** for visualization.
- While experimenting with multiple slicers, found the current layout to be the most informative and visually clear.

---

8. Categorising Issue Sizes

To analyse the relationship between **issue size** and **listing gain**, issue size groups were created using:

=IFS(A2<500,"0-500",A2<1000,"500-1000",A2<1500,"1000-1500",A2<2000,"1500-2000",

A2<2500,"2000-2500",A2<3000,"2500-3000",A2<3500,"3000-3500",A2<4000,"3500-4000",

A2<4500,"4000-4500",A2<=5000,"4500-5000",TRUE,"Above 5000")

- **Problem:** The resulting graph appeared **untidy** due to too many ranges.
- **Possible Solutions:**
    - Grouping gains into broader segments.
    - Using a lookup table with defined ranges.
- **Final Choice:** Continued using the IFS method since the dataset wasn’t very large.

---

9. Counting Number of IPOs

- Used a count formula within Pivot Table but found it disrupted the chart.
- Fixed it visually by **filling the count column with white color** to blend with the dashboard.
- Formula used:
- =XLOOKUP(A14,A14,GETPIVOTDATA("Count of IPO NAME",$A$3))
-

---

10. Counting Positive and Negative IPOs

Method 1: Filter Approach

- Created another pivot table and applied filter connection using a "greater than" group (>0, <0).
- Added "0" in the listing gains group to make formula handling easier.

Positive IPOs Formula

```
=SUMPRODUCT(
SUBTOTAL(103, OFFSET($B$4, ROW($B$4:$B$500)-ROW($B$4), 0, 1)) *
--(ISNUMBER($B$4:$B$500)) *
--($B$4:$B$500 >= 0)
)
```

Negative IPOs Formula

```
=SUMPRODUCT(
SUBTOTAL(103, OFFSET($B$4, ROW($B$4:$B$306)-ROW($B$4), 0, 1)) *
--(ISNUMBER($B$4:$B$306)) *
--($B$4:$B$306 < 0)
) - 1
```

- **Error Identified:**
    - Initially used **current gains** instead of **listing gains**, which broke the dashboard.
    - Fixed by correcting the data reference.
- **Result:**

- Negative companies = 226
- Positive companies = 225
- On deeper analysis, found **duplicate entries** (e.g., *Sai Silks (Kalaman)* listed twice with different QIB/NIB values).
- Removed duplicates to correct count accuracy.

---

11. Data Verification

- Verified duplicates using Excel's **Remove Duplicates** tool — no identical entries found except variations in QIB/NIB columns.
- Example duplicate:

Date	Year	Company	Issue Size	Group	Listing Gain	CMP	Status
27-Sep-23	2023	Sai Silks (Kalaman)	1201	1000–1500	12.35	186.3	>0
27-Sep-23	2023	Sai Silks (Kalaman)	1201	1000–1500	0	186.3	>0

- **Conclusion:** These minor variations caused count mismatches.

---

12. Fixing Count and Pivot References

- Adjusted the negative gain formula to remove 1 subtraction as it caused undercounting:
- =SUMPRODUCT(
- SUBTOTAL(103, OFFSET($B$4, ROW($B$4:$B$306)-ROW($B$4), 0, 1)) *
- --(ISNUMBER($B$4:$B$306)) *
- --($B$4:$B$306 < 0)
-)
-
- Moved total calculation to a separate sheet for stability.
- Used:
- =XLOOKUP('FOR NO OF +'!A11,'FOR NO OF +'!A11,GETPIVOTDATA("Count of IPO NAME",'FOR NO OF +'!$A$3))
-

to retrieve total counts dynamically.

---

13. Visualisation Adjustments

- Improved aesthetics by reducing decimals in the graph using MAX(0, …) wrapper:
- =MAX(0,
- SUMPRODUCT(
- SUBTOTAL(103,OFFSET(Total_vs_Gains!$C$4,ROW($C$13:$C$306)-ROW(Total_vs_Gains!$C$4),0,1))*
- --(ISNUMBER($C$13:$C$306))*
- --($C$13:$C$306<0)
-))
-
- Fixed reference error from C13 to C4.
- Removed decimals to make visuals cleaner.

---

14. Dashboard Development

- Created the **main dashboard** for IPOs (2021–2025).
- Collected metrics:
    - Total Subscription vs. Listing Gain
    - QIB vs. RII Subscription
    - Issue Size vs. Subscription
    - Negative Listing Gain vs. CMP
- Used multiple slicers for interactivity and visual storytelling.

---

15. Analytical Findings

1. Total Subscription vs. Listing Gains

- Example: *Patwari* IPO → **Subscription: 358.73x**, **Listing Gain: 39.02%**
- Indicates **strong correlation** between subscription demand and listing performance.

2. Listing Gain vs. Current Gain

- *Patwari:* Listing Gain: **39.02%**, Current Gain: **21.72%**

→ Suggests post-listing correction but sustained profitability.

3. QIB vs. RII Subscription

- *Patwari:* QIB = **190.97x**, RII = **57.75x**
- Institutional (QIB) investors show **stronger conviction** than retail (RII) investors.

4. Issue Size vs. Subscription

- *JSW Cement:* Issue Size = **3500–4000 Cr**, Subscription = **10.97x**
- Larger issues have **lower multiples** but still strong absolute capital inflow.

5. Negative Listing Gain vs. CMP

- *Bluestone:* Listing Gain = **–1.58%**, CMP = **₹546.35**
- Further calculation of **current gain = (CMP/Offer Price – 1)** would complete this analysis.

---

16. Outlier Detection

- Identified outliers using slicers and Pivot analysis.
- Most outliers were caused by categories with **only one IPO** entry in that range.

---

17. Final Observations

- Understanding the **purpose of analysis before cleaning and modeling** saves significant rework time.
- Functions like **IFS**, **XLOOKUP**, and **SUBTOTAL with OFFSET** proved extremely useful for dynamic analysis.
- The **Power Query function method** was crucial for real-time web scraping.

---

18. Key Learnings

1. Power Query can automate repetitive web scraping but needs careful parameter design.
2. Text inconsistencies (e.g., “Ltd.” vs “Limited”) heavily affect lookup accuracy.
3. CMP matching remains an open challenge — fuzzy lookup or APIs could solve it.
4. Graph readability improves with reduced decimals and grouped data ranges.
5. Cleaning duplicates early prevents count mismatches later in analysis.

19. Project Constraints

1. **Additional Key Elements for Analysis**
    - The project aimed to include other critical financial indicators such as **Profit**, **Revenue**, and **P/E (Pricing-to-Earnings) Ratio** for a more comprehensive evaluation of company performance.
    - These metrics were intended to complement the existing data and provide a deeper understanding of overall market health.

2. **Incomplete Dataset**
    - A **lack of complete and reliable data** for several of these financial parameters limited the scope of analysis.
    - Due to this data unavailability, the project remains **incomplete** in terms of holistic financial assessment.
3. **CMP (Current Market Price) Analysis Issue**
    - The **Current Market Price** of several companies could not be accurately analyzed.
    - This was primarily due to **limited knowledge on retrieving the correct stock names** aligned with their **respective market prices** from authentic sources.
    - As a result, **CMP-based comparisons** were excluded from the final evaluation.