

# Data Cleaning & Preparation Summary (Excel)

- Removed unnecessary rows from the Items and Price columns by applying filters to identify and delete rows containing blanks, errors, and “unknown” values.
- Replaced all occurrences of "error" and "unknown" with blank cells using Find and Replace.
- Standardized item pricing using the following reference values:
  - cookie = 1
  - tea = 1.5
  - coffee = 2
  - cake, juice = 3
  - sandwich, smoothie = 4
  - salad = 5
- Filled missing values in the Items and Price columns using the above reference table.
- Filled missing values in the Quantity column with the average value for each item group (average quantity = 3).
- Calculated the Total Spent column using the formula:  
 $=quantity*price$
- Filled blank cells in the Payment Method and Location columns by using Ctrl + G → Blanks → OK, typing = in the first selected cell, and pressing Ctrl + Enter to copy the value from the cell above.
- Added random dates to blank cells in the Transaction Date column using the formula:  
 $=DATE(2023,1,1) +RANDBETWEEN(0, DATE(2023,12,31)-DATE(2023,1,1))$

Store_ID	Area_Name	State
1005202	Nehru Place	Delhi
1005335	GK 1	Delhi
1005337	Korum Mall, Thane	Maharashtra
1005441	Ambience Mall, Gurugram	Haryana
1005678	Juhu Bawa, Mumbai	Maharashtra
1005757	Viviana Mall, Thane	Maharashtra
1005998	Koregaon Park, Pune	Maharashtra
1006041	Bandra West, Mumbai	Maharashtra
1006078	Amanora, Pune	Maharashtra
1006242	Backstage Building, Mumbai	Maharashtra
1006243	DLF Cyber Hub, Gurugram	Haryana
1006293	Phoenix Marketcity, Pune	Maharashtra
1006313	Netaji Subhash Place	Delhi
1006430	Aundh, Pune	Maharashtra
1006562	Whitefield, Bengaluru	Karnataka
1006687	Church Street, Bengaluru	Karnataka
1006696	Brigade Orion, Bengaluru	Karnataka
1006697	Mahavir Nagar, Mumbai	Maharashtra
1006818	High Street Phoenix, Lower Parel	Maharashtra
1006821	Inorbit Mall, Vashi	Maharashtra
1006847	EON Tech Park, Pune	Maharashtra
1006850	Infiniti Mall, Andheri	Maharashtra
1006859	Nirmal Building, Mumbai	Maharashtra
1006862	GK 2	Delhi
1007438	Forum Sujana Mall, Hyderabad	Telangana
1010018	Kokilaben Hospital, Mumbai	Maharashtra
1010298	Phoenix MarketCity, Mumbai	Maharashtra
1011485	Anna Nagar, Chennai	Tamil Nadu
1011504	Express Avenue Mall, Chennai	Tamil Nadu
1012310	Bergamo, Chennai	Tamil Nadu

1013108	Alwarpet, Chennai	Tamil Nadu
1013432	Inorbit Mall, Malad	Maharashtra
1013433	One BKC, Mumbai	Maharashtra
1013604	Gardens Galleria, Noida	Uttar Pradesh
1015283	One Horizon Center, Gurugram	Haryana
1015284	T3 Domestic Arrivals, Delhi Airport	Delhi
1016345	Sangam Solitaire, Pune	Maharashtra
1016965	T2 International Departure, Mumbai Airport	Maharashtra
1017071	Growels Mall, Mumbai	Maharashtra
1017119	Alcazar Building, Hyderabad	Telangana
1017223	Nesco Complex, Mumbai	Maharashtra
1017476	Candor TechSpace, Gurugram	Haryana
1017621	Silver Springs Building, Mumbai	Maharashtra
1017661	Kemps Corner, Mumbai	Maharashtra
1018000	Seawoods Grand Central, Navi Mumbai	Maharashtra
1018200	Victoria Mills, Mumbai	Maharashtra
1019329	Yamuna Expressway, Greater Noida	Uttar Pradesh
1019368	Khan Market	Delhi
1020122	32nd Milestone, Gurugram	Haryana
1020447	Ardee Mall, Gurugram	Haryana
1020494	WorldMark Aerocity	Delhi
1020801	Runwal Greens, Mumbai	Maharashtra
1020915	Green Park	Delhi
1021487	South City Mall, Kolkata	West Bengal
1021492	Commerz II, Mumbai	Maharashtra
1021694	Park Mansion, Kolkata	West Bengal
1021697	Pavillion Mall, Pune	Maharashtra
1021699	VR Chennai	Tamil Nadu

(from here the above table is used to add column cafe\_id,cafe\_location,state.)

- Added a new column named cafe\_id and assigned random cafe IDs using the formula:  
`=INDEX({1005202,1005335,1005337,1005441,1005678,1005757,1005998,1006041,1006078,1006242,1006243,1006293,1006313,1006430,1006562,1006687,1006696,1006697,1006818,1006821,1006847,1006850,1006859,1006862,1007438,1010018,1010298,1011485,1011504,1012310,1013108,1013432,1013433,1013604,1015283,1015284,1016345,1016965,1017071,1017119,1017223,1017476,1017621,1017661,1018000,1018200,1019329,1019368,1020122,1020447,1020494,1020801,1020915,1021487,1021492,1021694,1021697,1021699}, RANDBETWEEN(1,58))`
- Added a cafe\_location column using the lookup formula:  
`=LOOKUP(A2,N3:N61,O3:O61)`
- Added a final column named state using the formula:  
`=LOOKUP(A2,N3:N61,P3:P61)`

(Formulas may not be same for rows and columns so use it carefully to get the expected result.)

 After completing these steps, the dataset was fully cleaned, structured, and ready for power bi analysis.