Healthcare Claims Analysis- Where Is the Money Going?

Data Overview:

The dataset contains healthcare insurance claims data including member details, claim types, CPT and ICD codes, billed and paid amounts. SQL was used to clean the raw data by handling null values, removing duplicates, and standardizing fields. Aggregation queries were created to calculate total paid amounts, average paid per claim, and billed vs paid ratios for downstream analysis and dashboarding.

SQL (MySQL) Part — ***Data Preparation & Logic***

Show how raw data was transformed into analysis-ready data.

What to present :

- ✓ Data understanding (tables, rows, issues)
- ✓ Cleaning & transformation logic
- ✓ Business rules applied
- ✓ Final dataset output used in Tableau

Data source: https://www.analystbuilder.com/projects/healthcare-claims-where-is-the-money-going-TVHLQ

1. Database: MySQL
2. Tables used (claims, members)
3. Record count before cleaning

Data Cleaning :

- ✓ Check Duplicate/Null/Blank Values
- ✓ Create Duplicate Table (claims_staging & member_staging)
- ✓ Standardized date formats
- ✓ Data Type corrections
- ✓ Final clean Dataset (claims_clean & member_clean)

Data Analysing:

1. Claim Type Cost Breakdown
2. CPD & ICD Cost Drivers
3. CPT Drive Analysis
4. Member level Analysis
5. Billed vs Paid Ratio
6. Insurance Claims

SQL Logic (Key Queries):

1. Claim Type Cost Breakdown – **Inpatient** is **most expensive & Pharmacy least expensive** based on total paid amount.

2. CPD & ICD Cost Drivers

- Top 10 **CPT codes** by total paid amount.

cpt_code	Total_paid_amount	Claim_count	Avg_paid
85025	60	1	60.0000
31235	60	1	60.0000
99604	60	1	60.0000
90375	70	1	70.0000
99385	75	1	75.0000
81002	88	1	88.0000
43753	90	1	90.0000
50411	95	1	95.0000
9230	96	1	96.0000
83036	120	1	120.0000

- Top **10 ICD codes** by total paid amount.

icd_code	Total_paid_amount	Claim_count	Avg_paid
D12.7	120	1	120.0000
D50.6	50	1	50.0000
F32.9	168	1	168.0000
H60.3	145	1	145.0000
J06.9	68	1	68.0000
N40.1	150	1	150.0000
S52.101A	150	1	150.0000
Z20.822	70	1	70.0000
Z79.4	180	1	180.0000
Z99.89	150	1	150.0000

- CPT codes with a **high paid amount per claim**

cpt_code	Avg_paid_per_claim
99604	60.0000
85025	60.0000
31235	60.0000
90375	70.0000
99385	75.0000
99070	77.0000
99212	77.5000
81002	88.0000
43753	90.0000
50411	95.0000

3. Member – Level Analysis

member_id	claim_type	paid_amount_by_claim_type
1	Inpatient	28000
1	Outpatient	1600
1	Lab	400
1	Pharmacy	120
6	Inpatient	31500
6	Emergency	8000
6	Outpatient	2600
6	Lab	1200
8	Inpatient	17251
8	Emergency	4100
8	Outpatient	1341
8	Lab	280
8	Pharmacy	76
20	Inpatient	25050
20	Emergency	3000
20	Outpatient	1710
20	Lab	465
20	Pharmacy	150
28	Inpatient	25600
28	Outpatient	3600
28	Emergency	960
28	Lab	280
28	Pharmacy	120
32	Inpatient	35000

4. Billed vs. Paid Ratio

claim_type	provider_id	cpt_code	paid_ratio
Pharmacy	PRV00004	34567	1.0000
Pharmacy	PRV01234	11111	1.0000
Pharmacy	PRV00005	10345	1.0000
Lab	PRV04567	23456	1.0000
Pharmacy	PRV07890	34567	1.0000
Pharmacy	PRV04567	99070	1.0000
Outpatient	PRV12345	55210	1.0000
Lab	PRV11223	80050	1.0000
Emergency	PRV55667	99285	1.0000
Outpatient	PRV33449	93000	1.0000
Lab	PRV54321	1234	1.0000
Lab	PRV67890	80050	1.0000
Lab	PRV00007	78901	1.0000
Outpatient	PRV00001	50123	1.0000
Lab	PRV00003	50377	1.0000
Pharmacy	PRV00004	50411	1.0000
Lab	PRV01004	999	1.0000
Pharmacy	PRV01005	1	1.0000
Lab	PRV11111	11111	1.0000

5. Look for claim types, insurer pays significantly less or significantly more than the billed amount.

claim_type	Avg_Paid_Ratio	Payment_flag
Inpatient	0.7375	Underpaid
Emergency	0.7663	Underpaid
Outpatient	0.8030	Within Expected Range
Pharmacy	0.8895	Within Expected Range
Lab	0.9078	Within Expected Range

MySQL to Tableau : All data preparation was performed in MySQL. The finalized, cleaned tables were then imported into Tableau using extract connections for dashboard development."