**Clarification questions:**

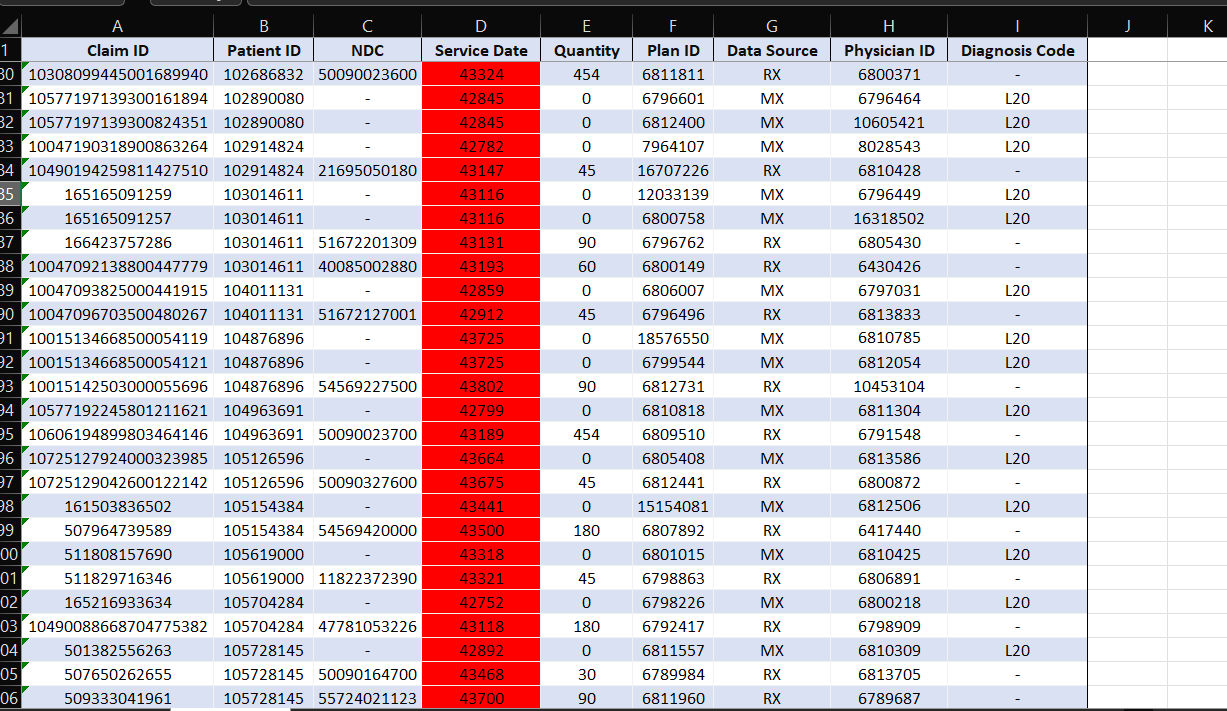
* What is the drug form of staquili ? Or Is it that the product consists of different types of drugs which are mentioned?
* Which metric are we tracking for considering top payer name?
* In Q6. what is the number of top payers required (Top1, Top 10, Top 100) ?
* What factors/metrics are required to track for understanding more imp payer channels?

**Q1. What types of Data checks you'll apply to the provided data to ensure the integrity and completeness of the given data tables?**

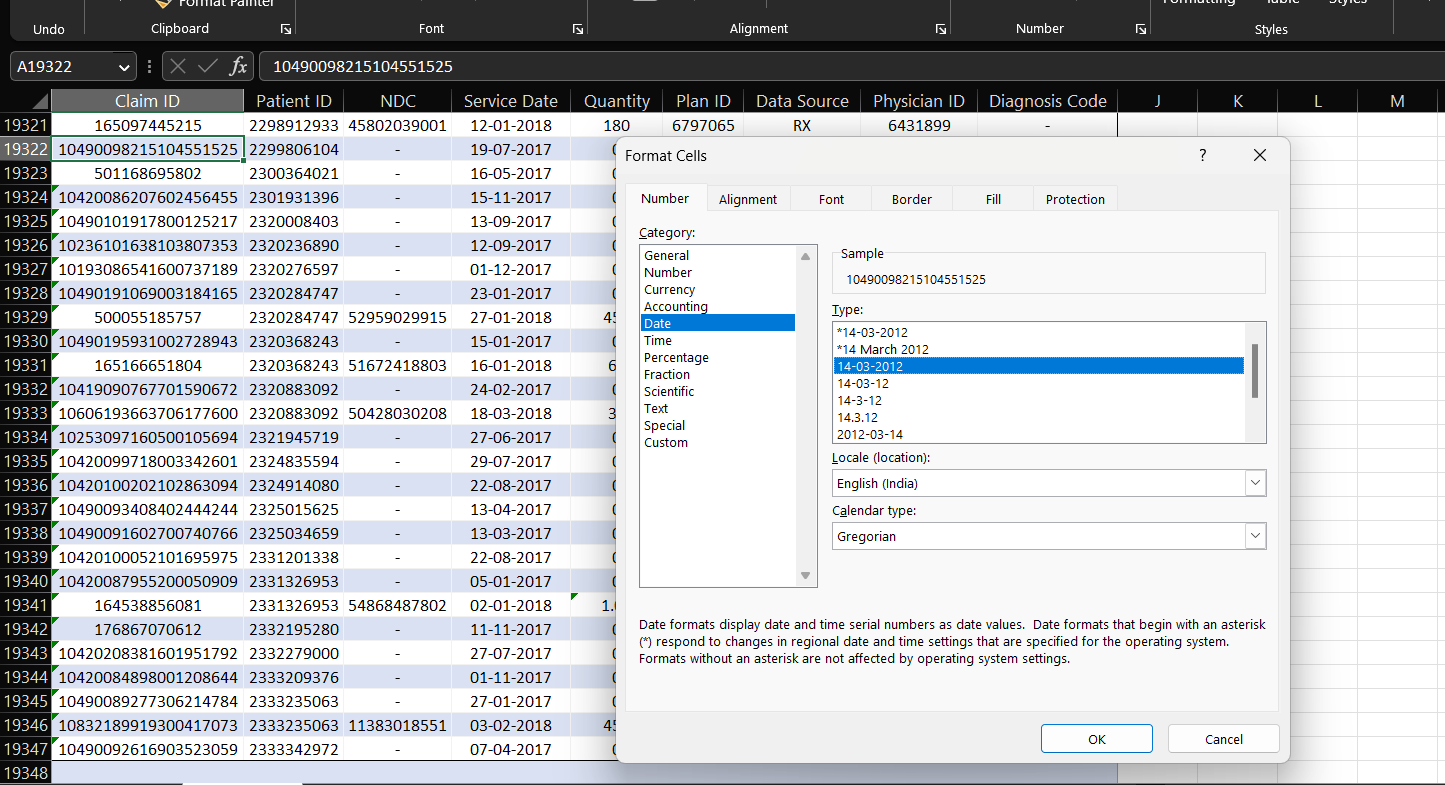
**SOLUTION:**

**1) Data Exploration Checks**

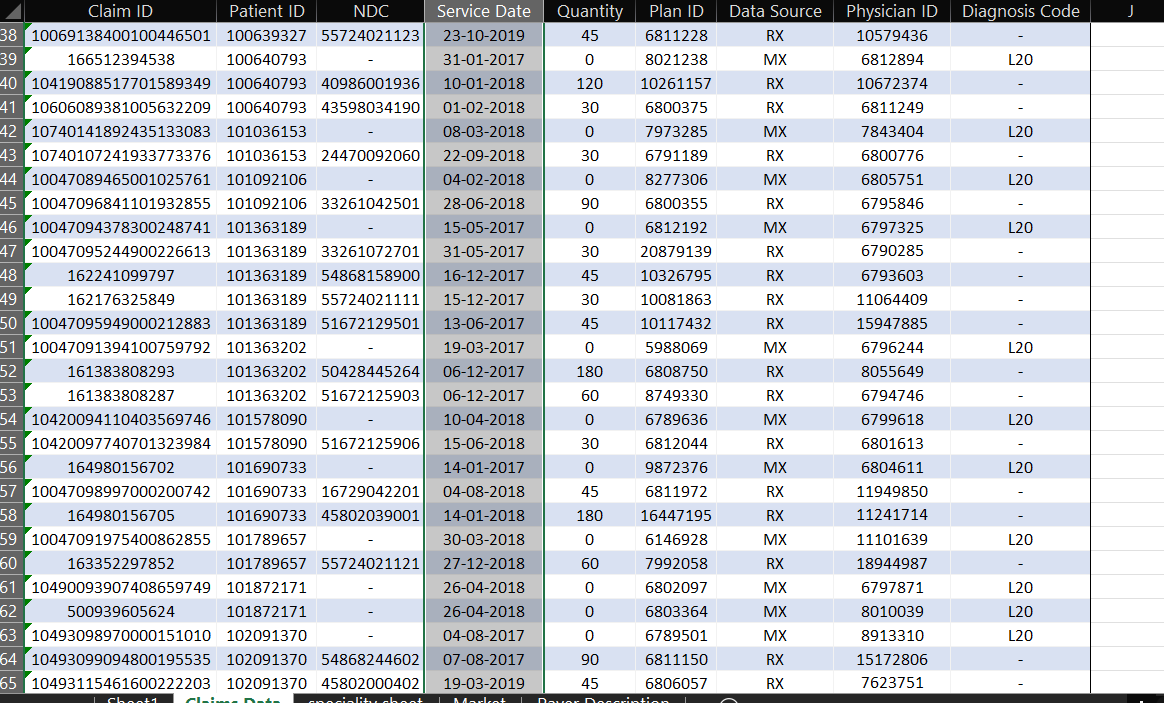
* By exploring the data we found I found that **Service data** in the **Claims** is not in the correct format.
* Hence converted the complete data in the correct(Date) format.



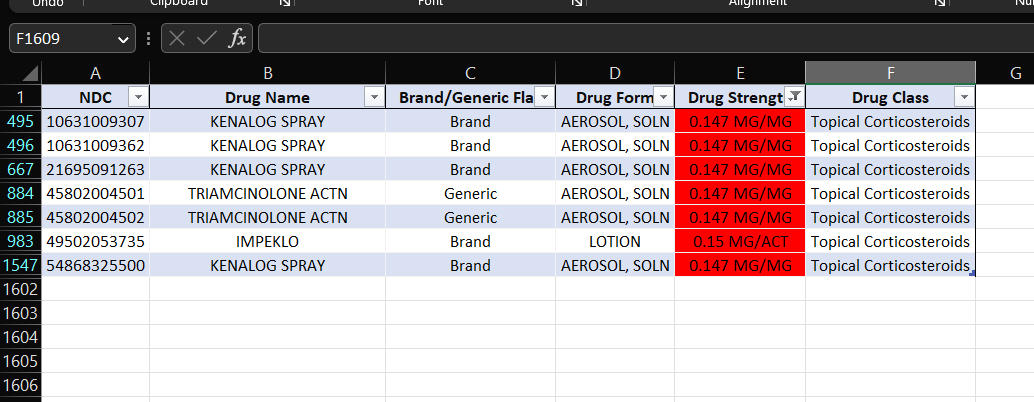
-Formatting is done to convert it into date format



-Final



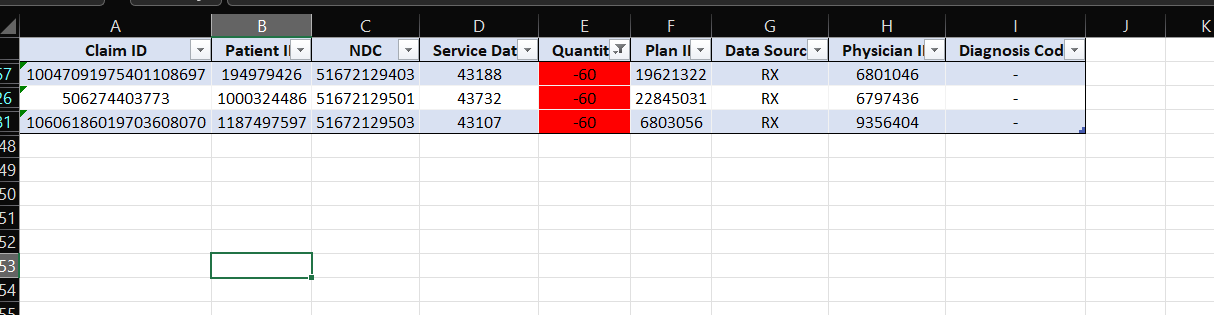
**-** In **Market Definition** table, the units for certain inputs is **MG/MG** and **MG/ACT,** while the majority of the data is in percentage. Hence we need to convert MG/MG and MG/ACT into percentage.



**2) Data Accuracy Test:**

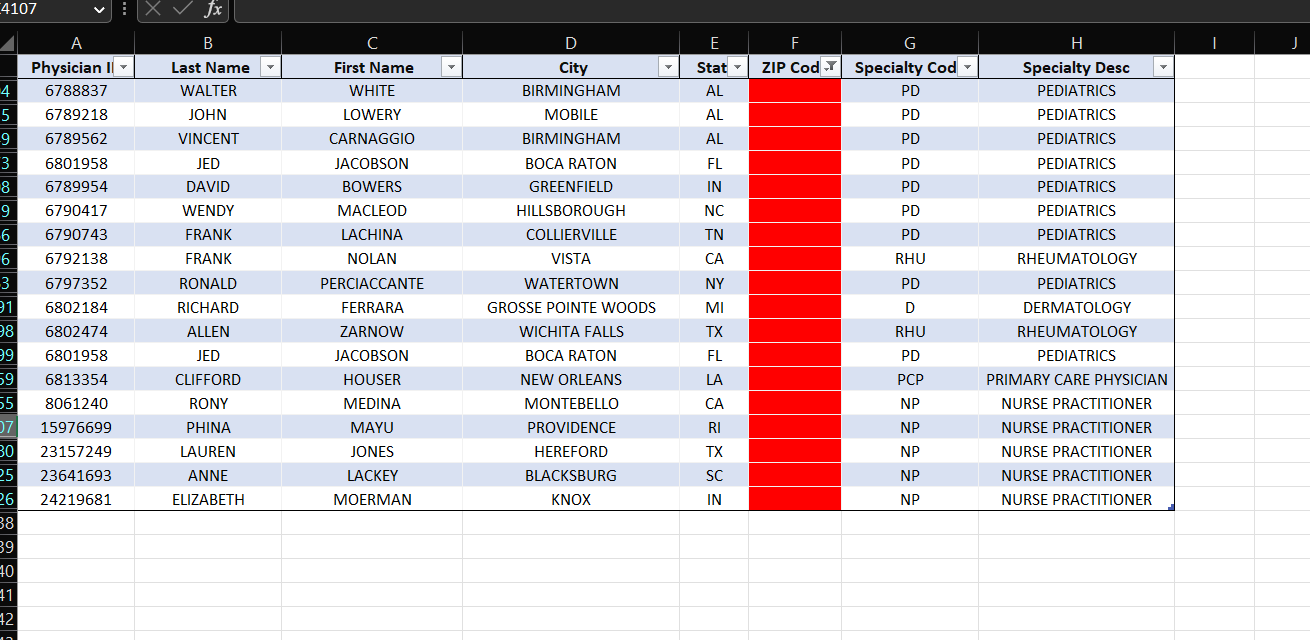
**Quantity** of the drug is negative for some entries in **Claims Data** which is not possible. Hence the data is wrong.

Deleted the rows from the data

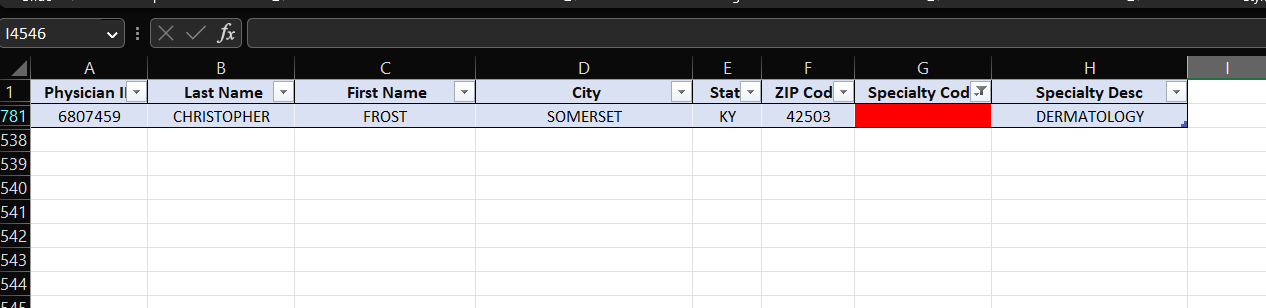


**3) Missing Data:**

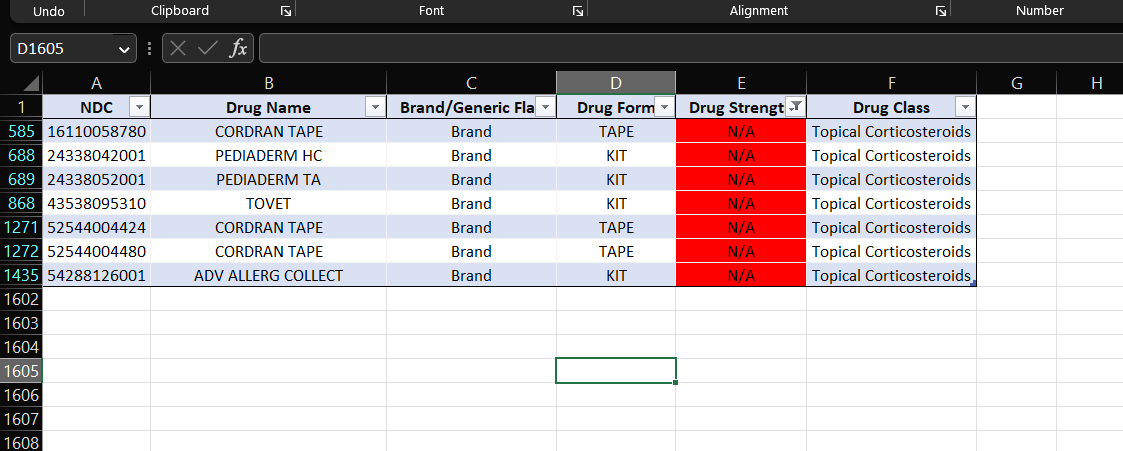
- The data of **Zip codes** of the regions is missing for some of the rows from **Specialty Description** table.



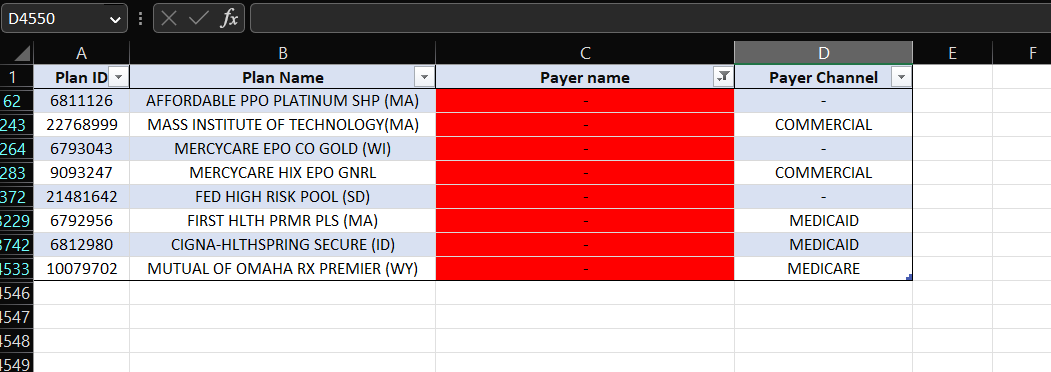
- The data of **Specialty Code** is missing for row in the **Specialty Description** table.



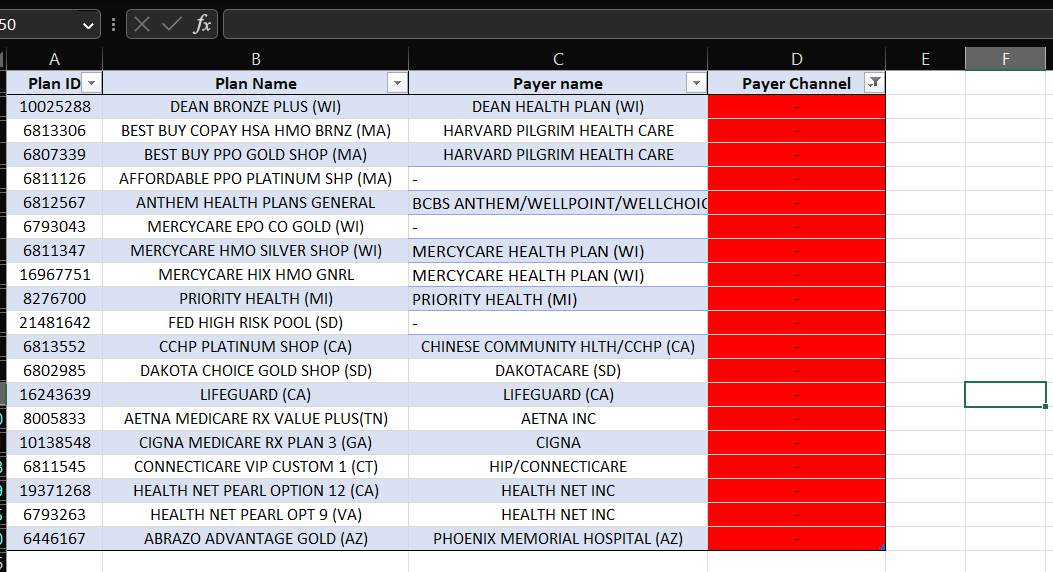
-In **Market Definition** table,the data of **Drug strength** is missing for some rows.



- Data of **Payers name** is missing in some of the entries in **Plan description.**



-Data of **Payer Channel** is missing for some of the rows in **Plan Description.**



**2) What % of patients move from diagnosis to treatment? What could be the probable reasons for some patients not getting on treatment?**

**Solution:**

No of people diagnosed: **7261** (There are few people who are diagnosed 2 times)

No of people started treatment: **4507**

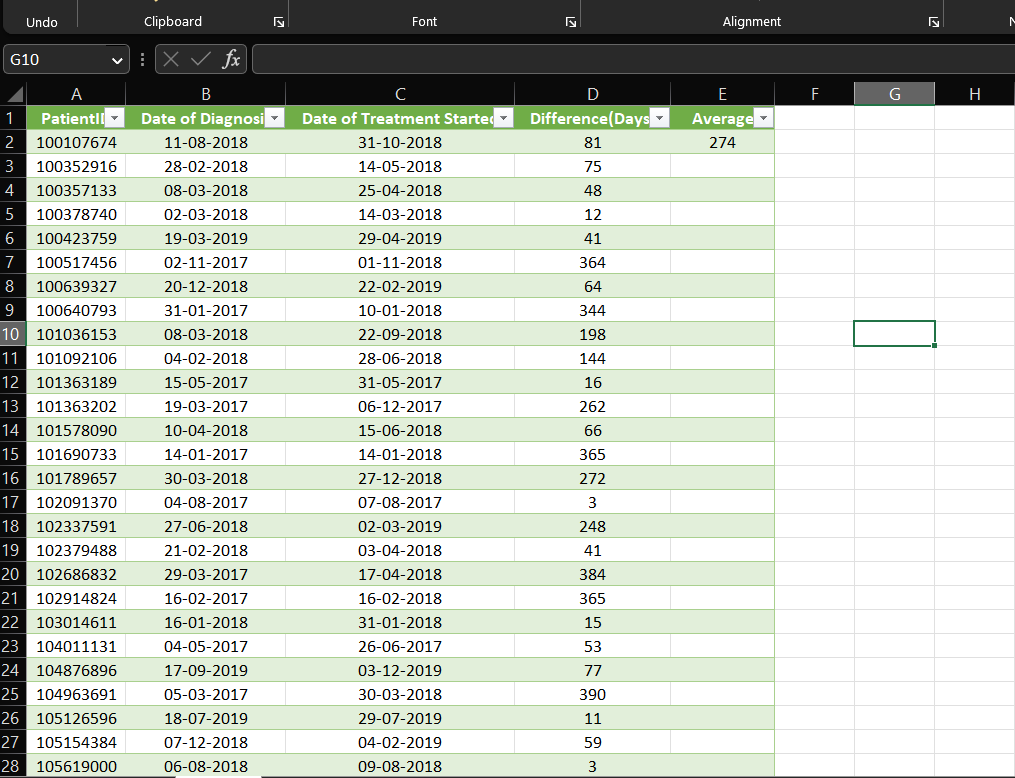
% of people moving from diagnosis to treatment: 4507\*100/7261= **62.071%.**

**Probable reasons for some patients not getting on treatment :**

* Financial barriers, such as lack of insurance or inability to afford the treatment.
* Lack of transportation to get to medical appointments.
* Limited access to healthcare providers in their area.
* Difficulty understanding or trusting the medical system.
* Fear or mistrust of the medical profession.
* Difficulty understanding the treatment or its potential side effects.
* Belief that the treatment is unnecessary or unproven.
* Co-occurring mental health or substance abuse issues.
* Cultural or linguistic barriers.
* Difficulty managing the treatment regimen.
* Personal beliefs or values that conflict with the treatment.

**3) On average, how long (in days) do patients take to move from diagnosis to treatment?**

**Solution:**

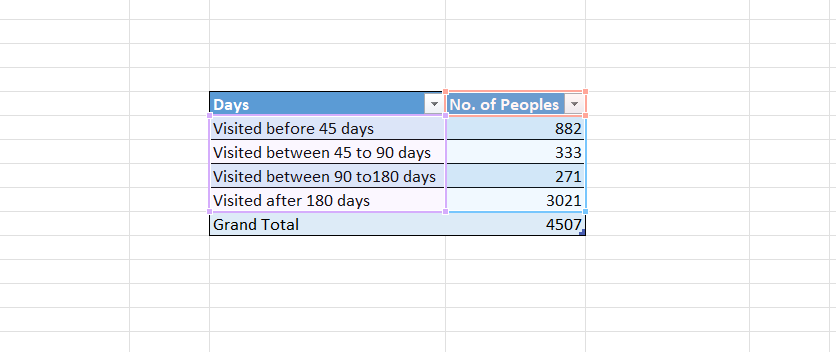
****

**On average patient takes 274 days to move from Diagnosis to treatment.**

**4) Create a graph to show patients started with the treatment within 45, 90, 180 and post 180 days.**

**Solution:**

**-**Table showing the required data

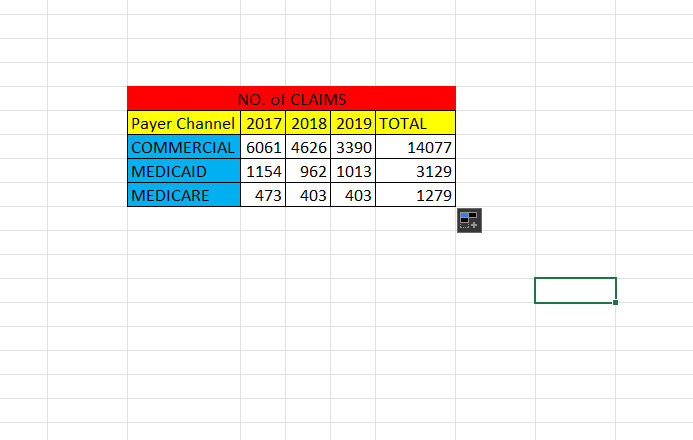


-Corresponding Graph

5) **What is the claim share of the payer channel over the years? Provide insights and visualizations. Do you see any change in claim pattern across years?**

**Solution:**

**-Claim share of Different payer channel over the years :**



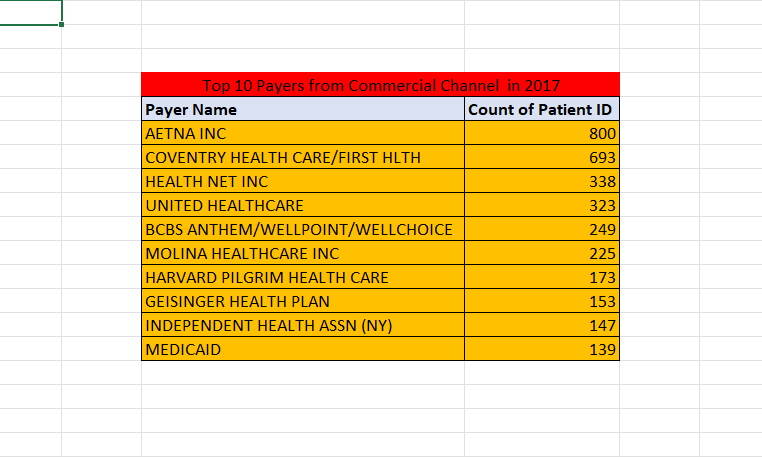
-**Insights and Visualization:**

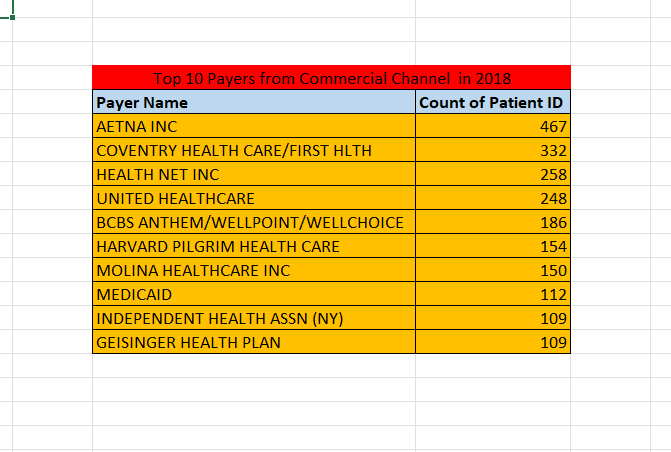
**- We can observe that no. of claims by Medicaid and Medicare payer channel remains throughout the consecutive years**

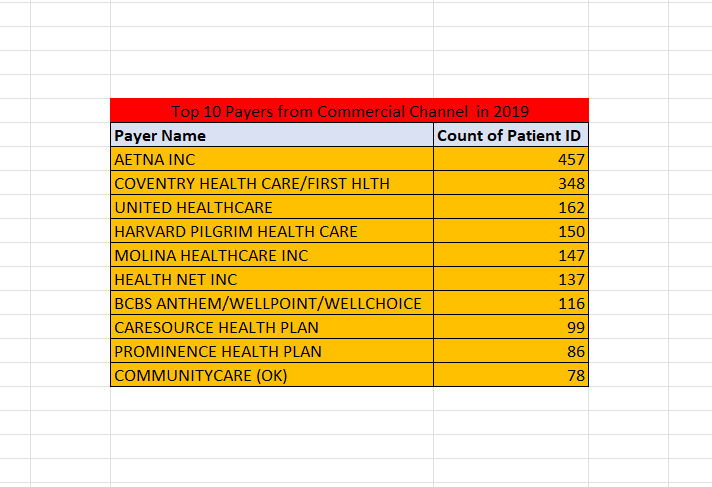
**But it is continuously decreasing for the Commercial Payer channel**

**6) Who are the top payers (Payer name) within the commercial channel over the years?**

**Solution:**







**7) Considering Dermatology and Nurse Practitioners specialty physicians as the major targets for ProcDNA, which payer channel becomes more important and why?**

**Solution:**

**Commercial Channel becomes very important because the count of this is the largest for Dermatologist and Nurse practitioners**

