**DRUG PREDICTION FOR DIFFERENT DISEASE BASED ON PATIENT’S HISTORICAL DATA**

**Business Problem:**

To do analysis based on diseases and drugs and the condition of the patients according to the seasons and to predict the major drugs used by the people and to Perform EDA and Analytics.

**Business Objective:**

1.Maximize prediction accuracy

2.Identify the seasonal diseases and general diseases

3.Maximize the frequently used drugs according to the diseases

4.Minimize health issues

5.Maximize profits

**Business Constraints:**

To promote high standards in pharmaceutical product production, quality assurance and

preparation of compound medicinal prescriptions, pharmaceutical and other appropriate products.

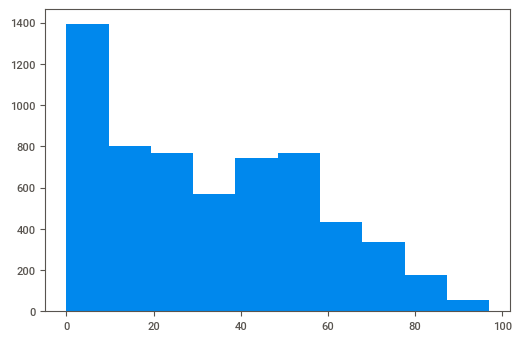
**Data Preprocessing :**

* No missing values in the dataset
* No duplicate values
* No outliers Found

**Exploratory Data Analysis insights:**

**Skewness**

* The data points lies on the right side and it is positively skewed, Hence it is positively distribuited Mean>Median>Mode.



* Visualized using Sweetviz auto library

**These are all the top 5 frequent condition of the patient**

Streptococcal sore throat (disorder) 1074

contraception 888

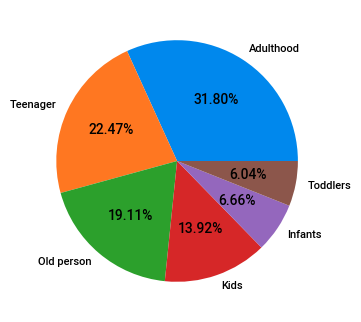
Acute bronchitis (disorder) 508

severe pain 470

Coronary Heart Disease 320

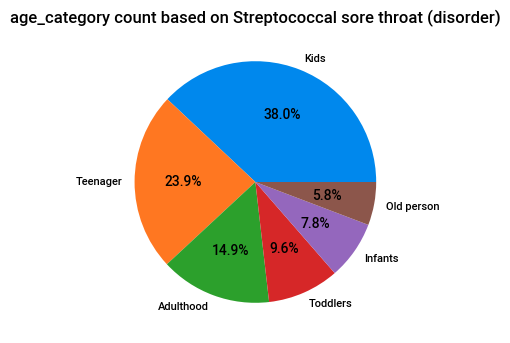
**Prediction about the age\_category and the specific Condition of Patient**

* For a single condition how patients suffer based on their age\_category.

In Age\_category we see that Adulthood is mostly affected of 31.80% ratherthan the other categories. 

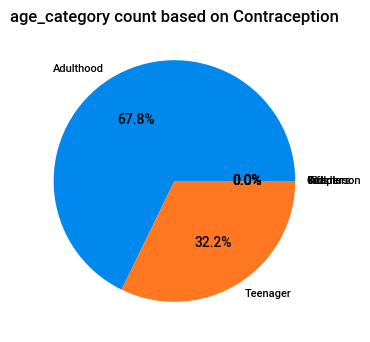
**Prediction about the age\_category and the specific Condition of Patient**

* In this pie chart we see that kids suffer more from Streptococcal sore throat (disorder) of 38% rather than others.

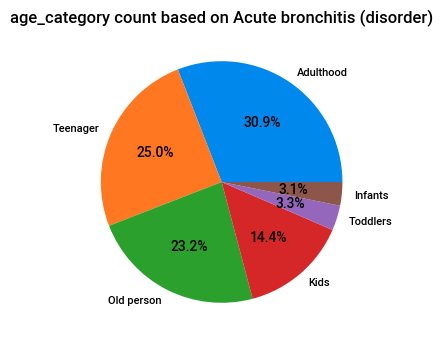


* the exact count is
  + Kids-38%
  + Teenager-23.9%
  + Adulthood-14.9%
  + Toddlers-9.6%
  + Infants-7.8%
  + Old person-5.8%
* In Contraception case Adulthood are mostly affected rather than the other age categories

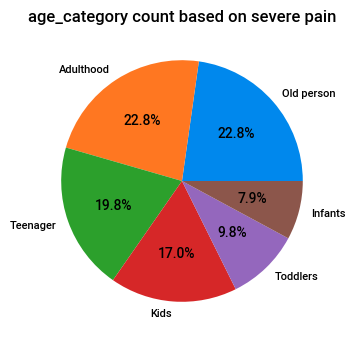
.



* Adulthood-67.8%
* Teenager-32.2%
* In the case of Acute bronchitis (disorder) Adulthood is majorly affected rather than the other age groups.

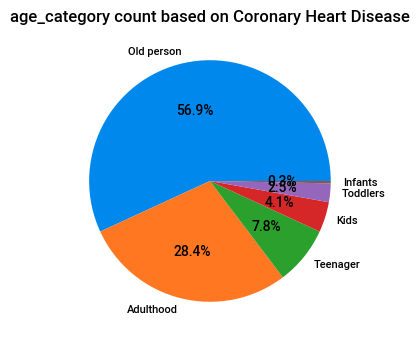


* The exact percentage is
* Adulthood-30.9%
* Teenager-25.0%
* Old person-23.2%
* Kids-14.4%
* Toddlers-3.3%
* Infants-3.1% this is the least that was affected.
* Adulthood and old person have 22.8% and they are equally affected in the condition of severe pain.



* The exact percentage is
* Adulthood-22.8%
* Old person-22.8%
* Teenager-19.8%
* Kids-17.0%
* Toddlers-9.8%
* Infants-7.9%

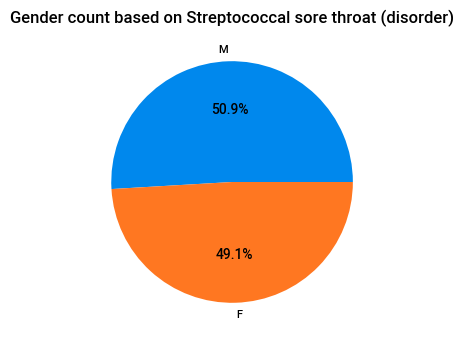
* In the case of coronary heart disease old persons is majorly affected followed by adulthood rather than the other age groups.



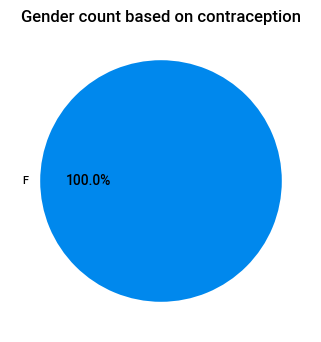
* The exact percentage is
* Old person-56.9%
* Adulthood-28.4%
* Teenager-7.8%
* Kids-4.1%
* Toddlers-2.5%
* Infants-0.2% this is the least that was affected.

**Prediction on Genders and the condition of the patients**

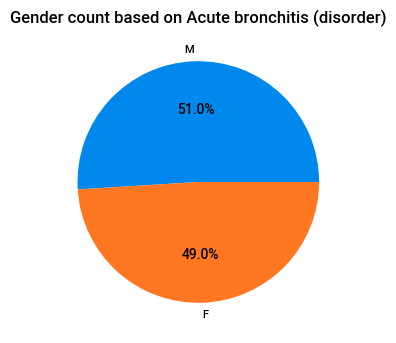
* In Streptococcal sore throat disorder Males are majorly affected when compared to females but there is no major differences males are of 50.9% and females are of 49.1%



* In the case of contraception disease Females are fully affected of 100% because this disease cannot cause in males

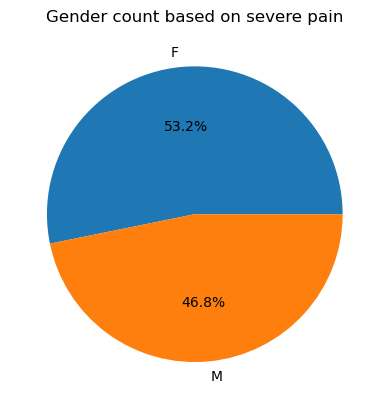


* In the case of Acute bronchitis (disorder) Males are majorly infected when compared with females but there is no major differences a slight variation is males are of 51% and females are of 49%



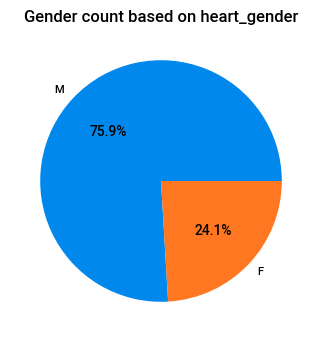
* In the case of severe pain females are majorly caused ones rather than Males.

for females the percentage is 53.2% and for males it is 46.8%



* The percentage rate is
* Males-75.9%
* Females-24.1%

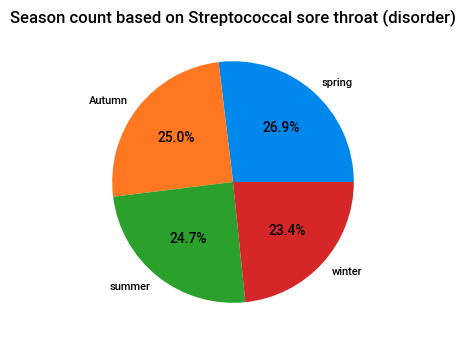
In the case of coronary heart disease Males are mostly affected than Females



* The percentage rate is
* Males-75.9%
* Females-24.1%

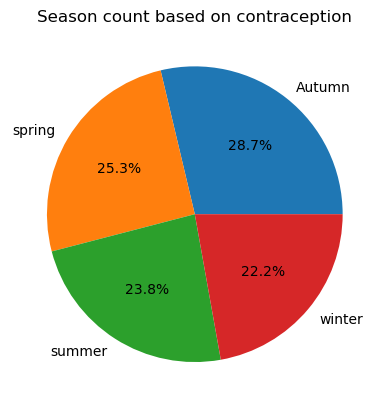
**Prediction according to Seasons and condition of the patients**

* In the spring season Streptococcal sore throat (disorder) was majorly affected by the people.



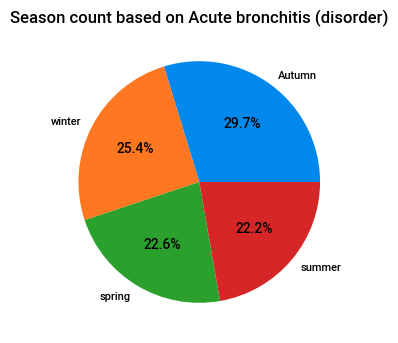
The exact percentage rate is

* Spring- 26.9%
* Autumn -25%
* summer-24.7%
* winter-23.4%
* In the Autumn season Contraception disease was majorly affected by females and hence it is a general disease.



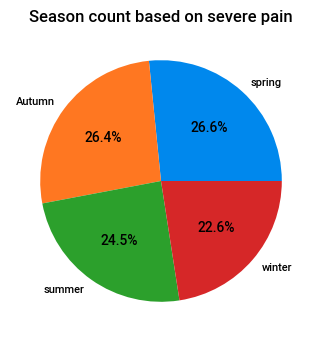
The exact percentage rate is

* Autumn -28.7%
* Spring- 25.3%
* summer-23.8%
* winter-22.2%
* In the Autumn season Acute bronchitis (disorder) disease was majorly affected by people and hence it is a general disease.



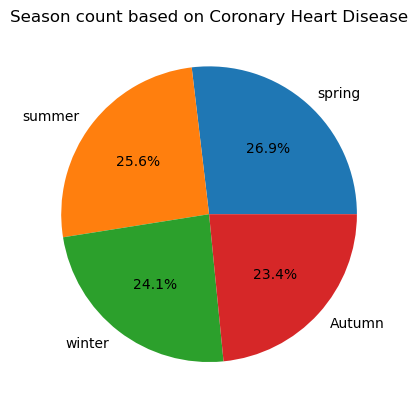
The exact percentage rate is

* Autumn -29.7%
* winter-25.4%
* Spring- 22.6%
* summer-22.2%
* In the spring season severe pain disease was majorly affected by people and hence it is a general disease.



The exact percentage rate is

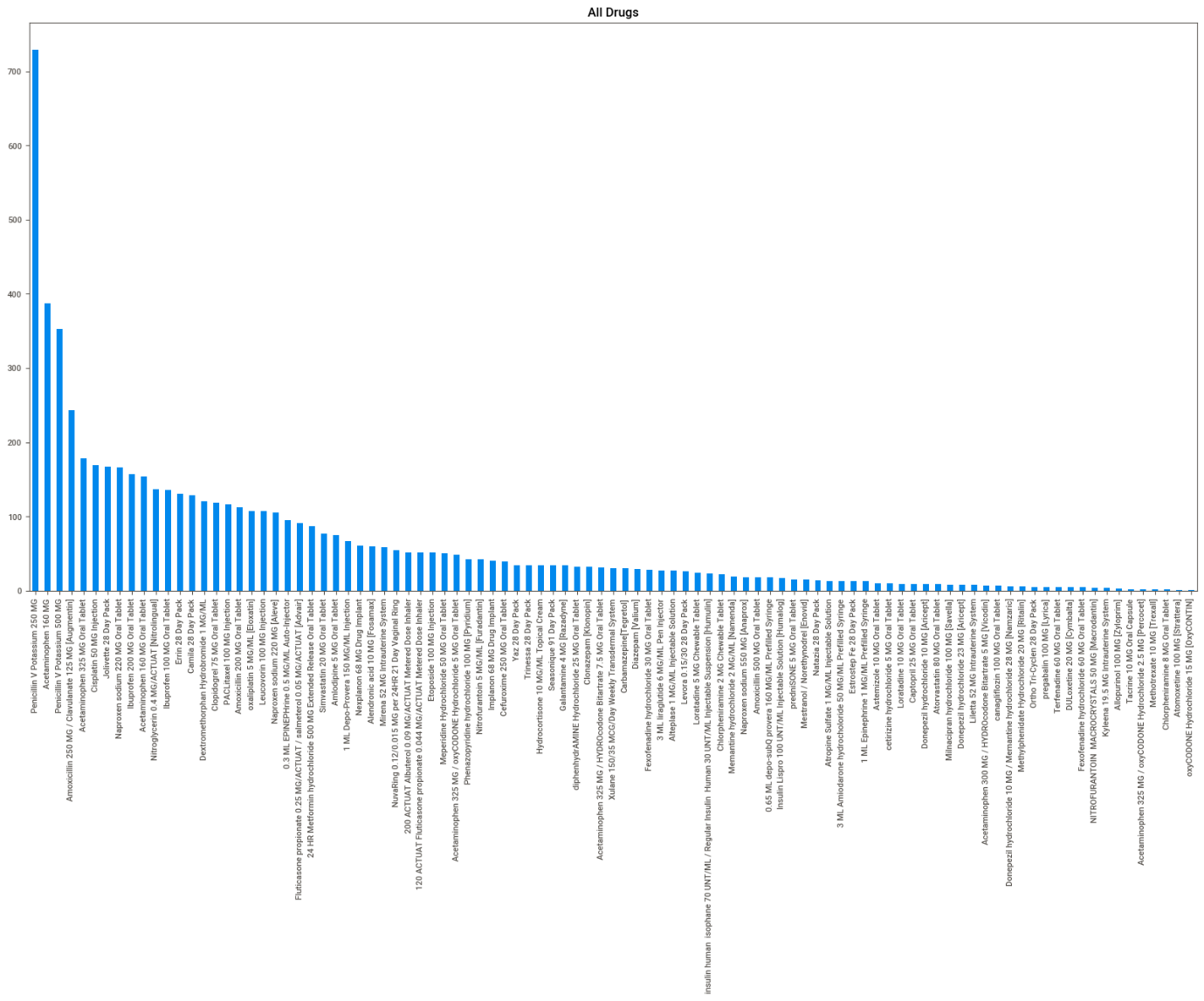
* Spring- 26.6%
* Autumn -26.4%
* summer-24.5%
* winter-22.6%
* In the Spring season Coronary Heart Disease disease was majorly affected by people and hence it is a general disease.



The exact percentage rate is

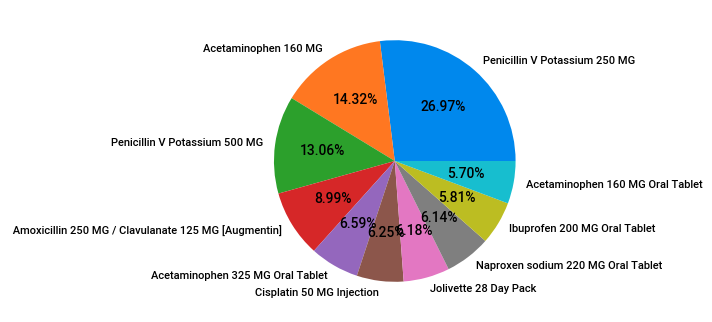
* Spring- 26.9%
* summer-25.6%
* Autumn -23.4%
* winter-24.1%

**The Quantity of each and every drug, visualized using matplotlib in descending order**

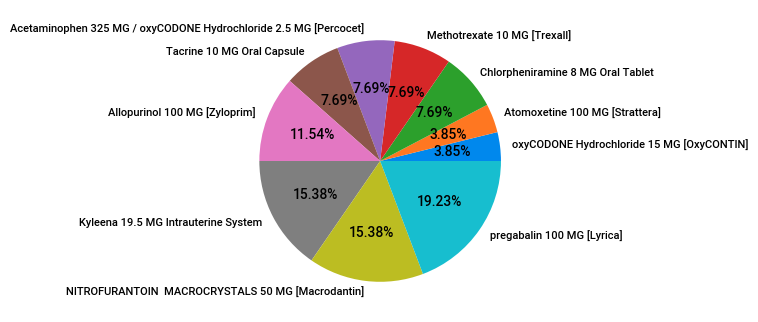


**These are the top 10 drugs identified and visualized using matplotlib**

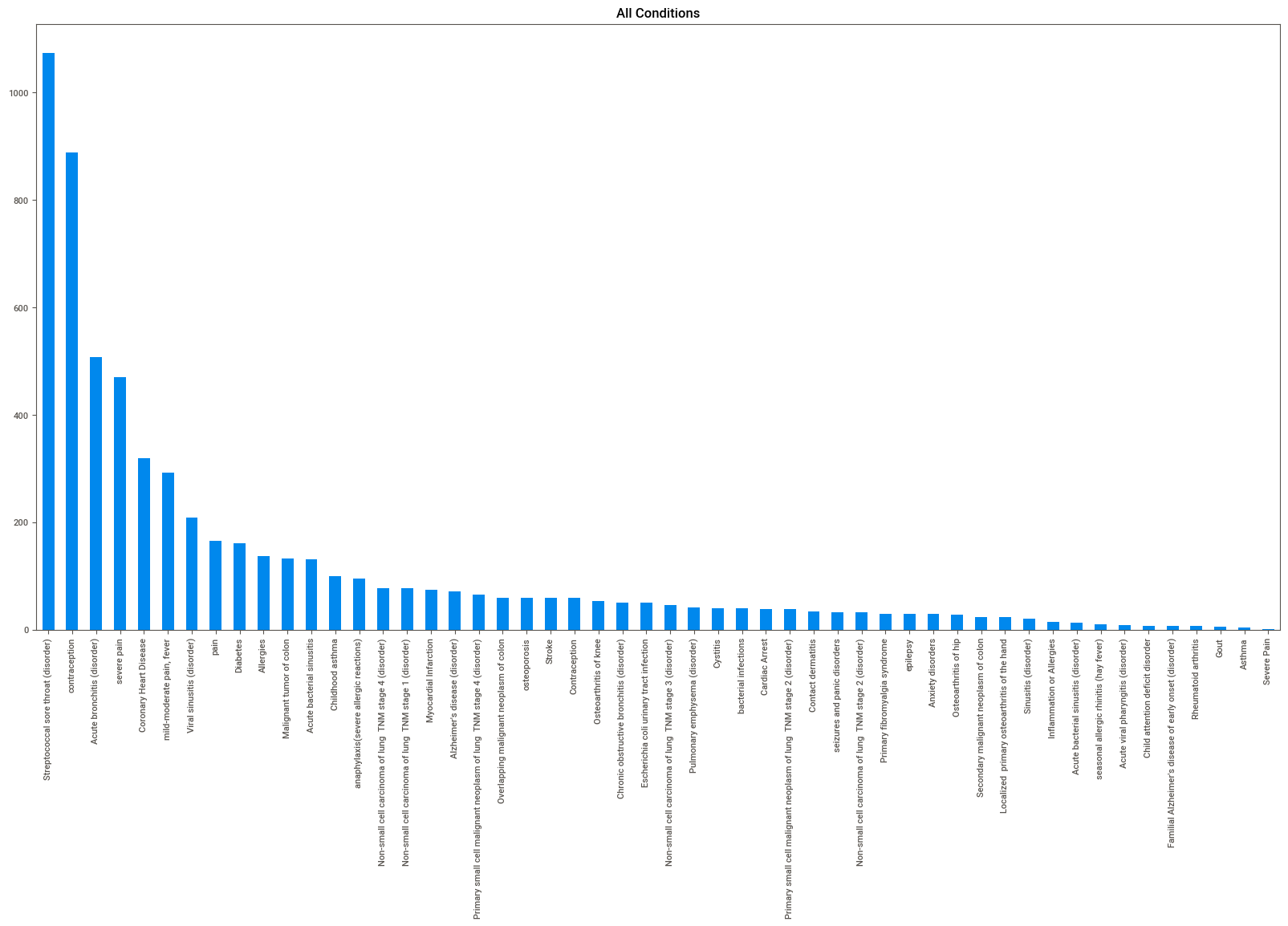
* In this Pencillin V potassium is majorly used



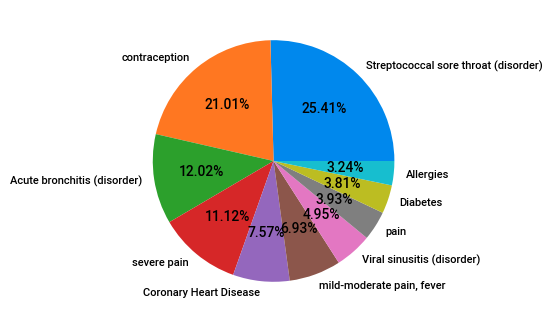
**These are the least 10 drugs identified and visualized using matplotlib**



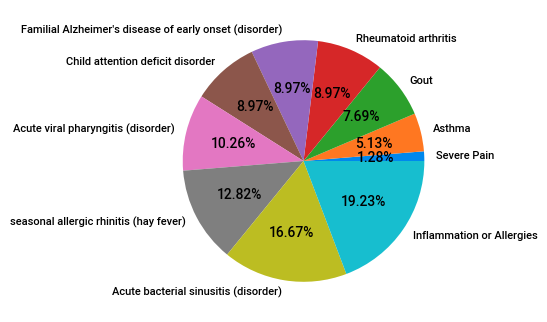
**Frequency of each and every condition of the patient is visualized using matplotlib in descending order**



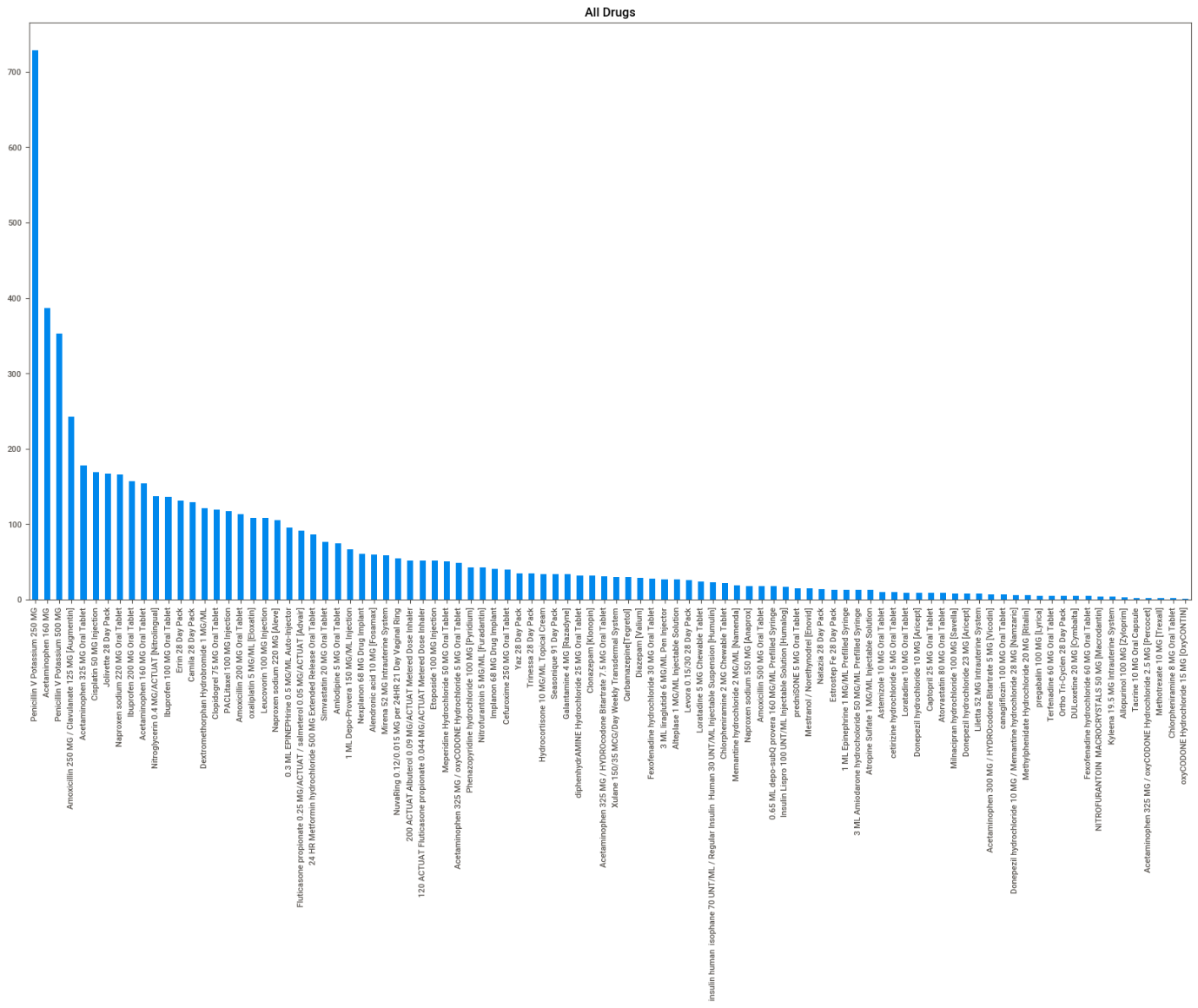
**These are the top 20 conditions identified and visualized using matplotlib**



**These are the Least 10 conditions identified and visualized using matplotlib**

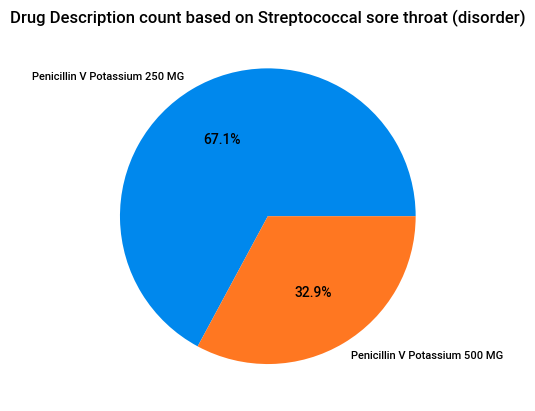


**Prediction according to Drug Description and condition of the patients**

****

These are the all drugs mentioned in the dataset

**Count of Streptococcal sore throat (disorder) based on Drug Description**



This pie chart represents the drug description based on Streptococcal sore throat (disorder)

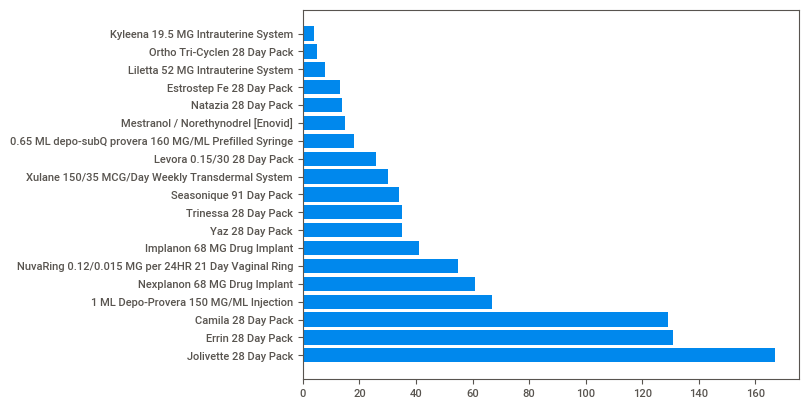
for this disease Penicillin V Potassium are majorly used

the exact percentage of MG in Penicillin V Potassium is

* Penicillin V Potassium 250MG - 67.1%
* Penicillin V Potassium 500MG - 32.9%

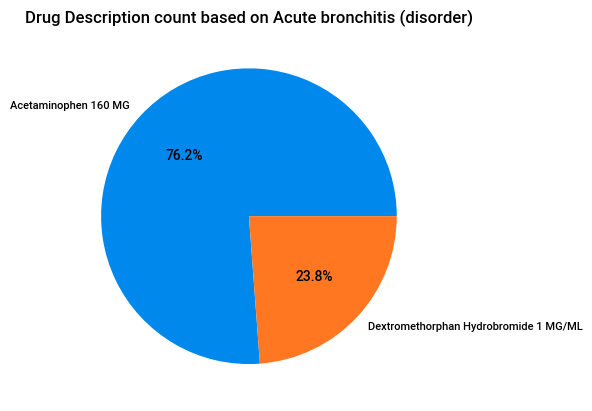
Here, Penicillin V Potassium 250MG dosage is often used.

**Count of contraception based on Drug Description**

****

In this pie chart it represents the drug description count based on contraception is shown in bar graph

**Count of Acute bronchitis (disorder) based on Drug Description**



This pie chart represents the drug description based on Acute bronchitis (disorder)

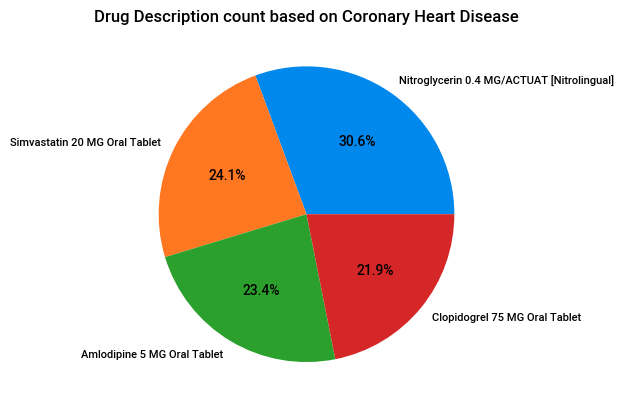
for this disease Acetaminophen 160 MG and Dextromethorphan Hydrobromide 1 MG/ML are majorly used

the exact percentage of MG in this medicine is

* Acetaminophen 160 MG - 76.2%
* Dextromethorphan Hydrobromide 1 MG/ML - 23.8%

Here, Acetaminophen 160 MG is often used for Acute bronchitis (disorder)

**Count of Coronary Heart Disease based on Drug Description**



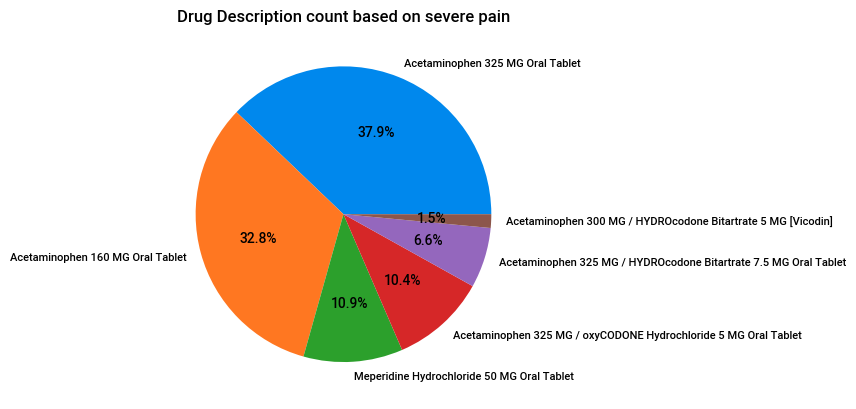
This pie chart represents the drug description based on Coronary Heart Disease

for this disease the drugs which were used is shown.

the exact percentage in this medicines for Coronary Heart Disease is

* Nitroglycerin 0.4 MG/ACTUAT [Nitrolingual] - 30.6%
* Simvastatin 20 MG Oral Tablet - 24.1%
* Amlodipine 5 MG Oral Tablet - 23.4%
* Clopidogrel 75 MG Oral Tablet - 21.9%

**Count of severe pain based on Drug Description**

****

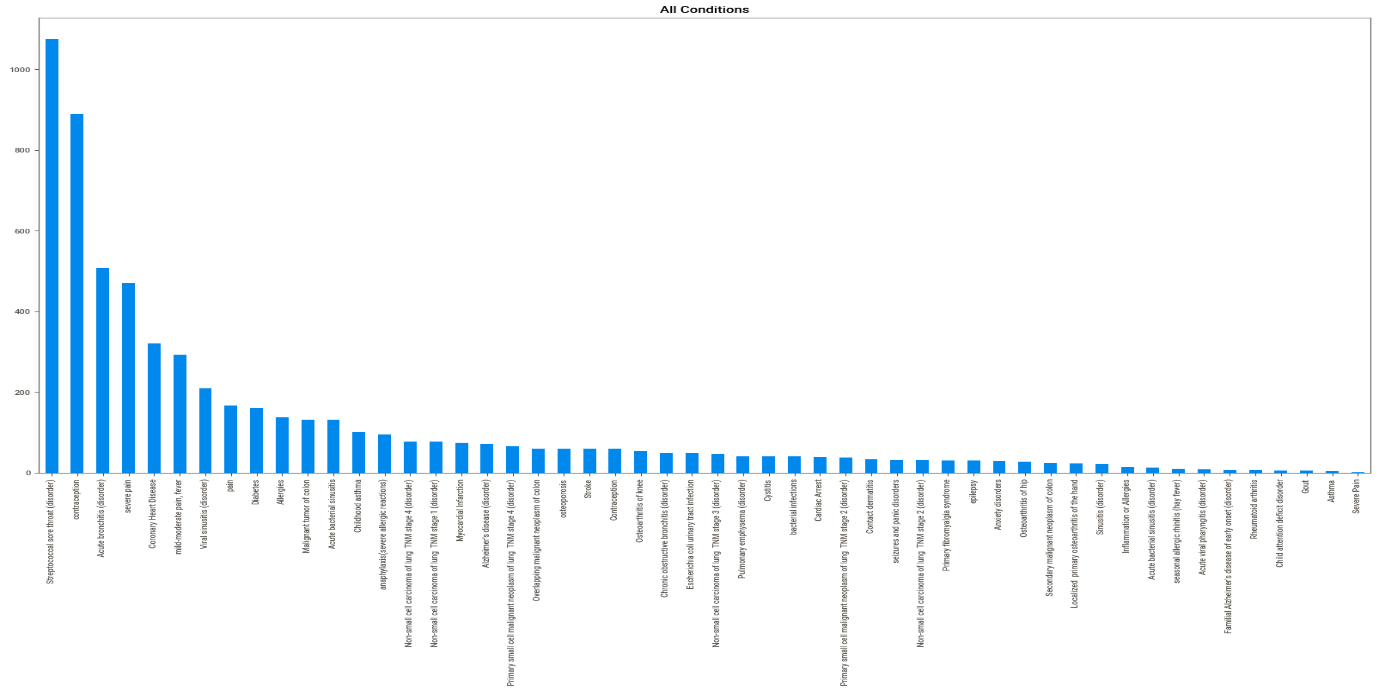
This pie chart represents the drug description based on severe pain

for this disease these drugs are used for treatment

the exact percentage in this medicine is

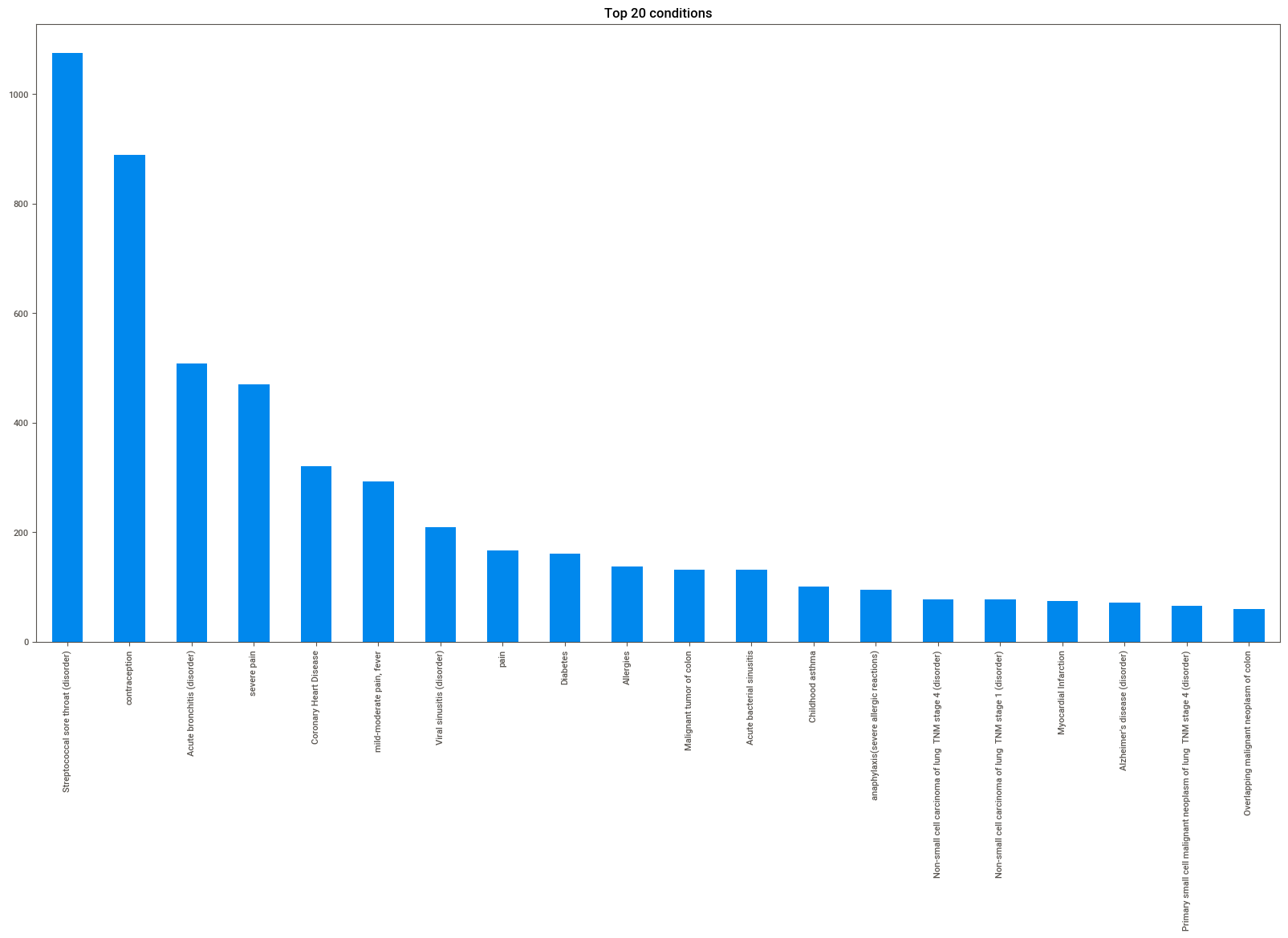
* Acetaminophen 325 MG Oral Tablet - 37.9%
* Acetaminophen 160 MG Oral Tablet - 32.8%
* Meperidine Hydrochloride 50 MG Oral Tablet - 10.9%
* Acetaminophen 325 MG / oxyCODONE Hydrochloride - 10.4%
* Acetaminophen 325 MG / HYDROcodone Bitartrate - 6.6%
* Acetaminophen 300 MG / HYDROcodone Bitartrate - 1.5%

**All conditions**

****

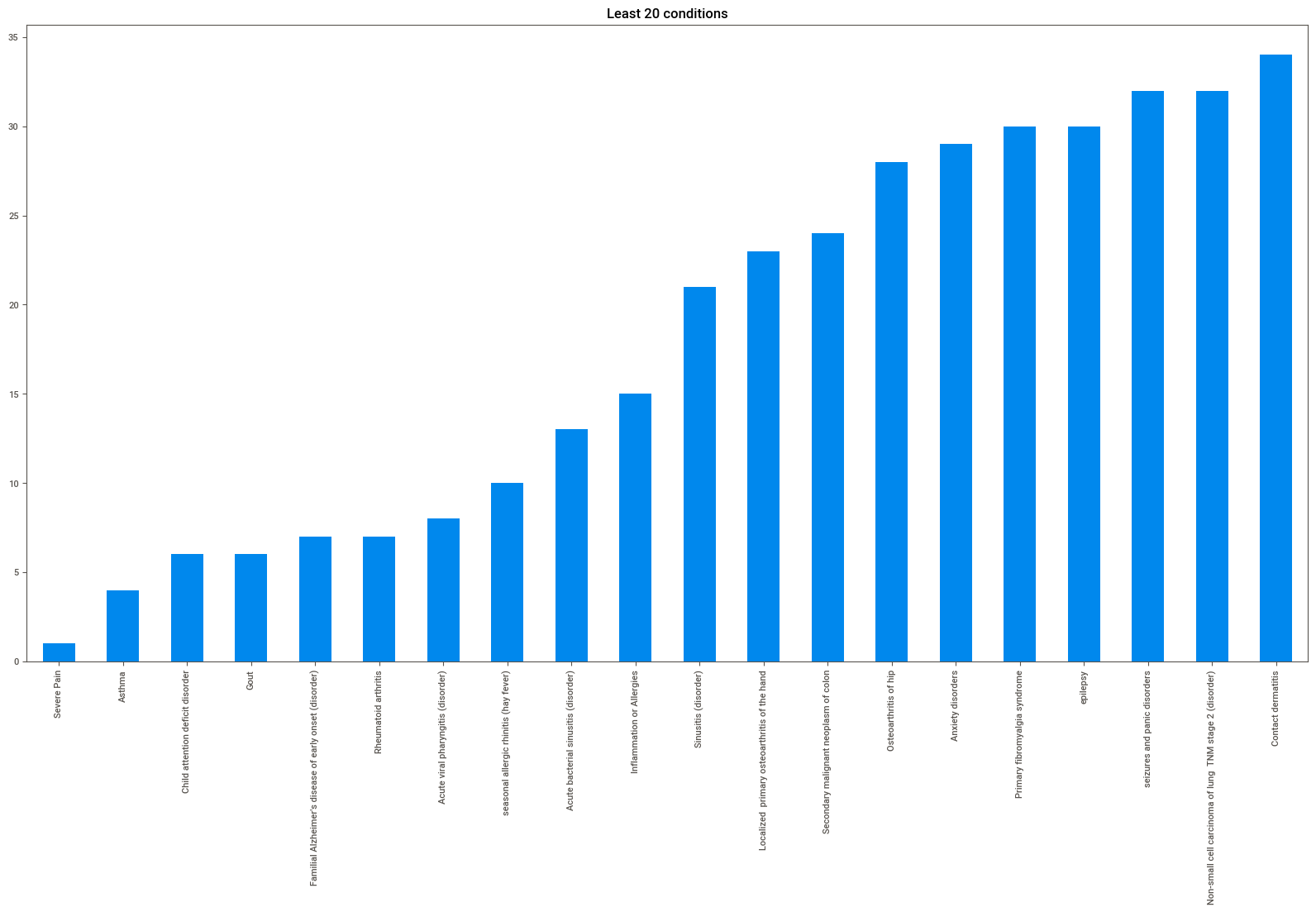
Frequency of each and every condition of the patient is visualized in descending order

**Top 20 conditions**

****

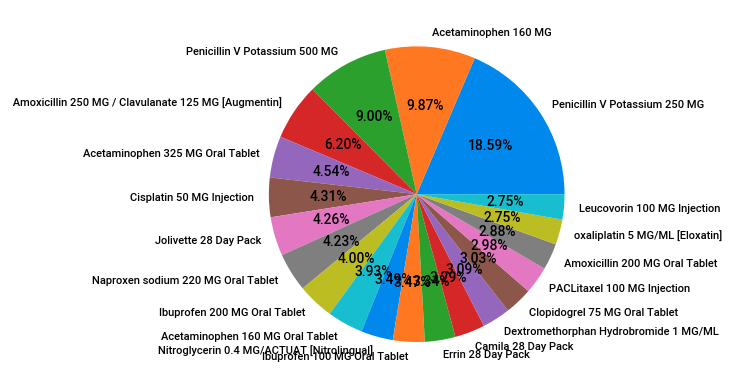
These are the top 20 conditions identified and visualized using bar graph

**least 20 conditions**

****

These are the Least 20 conditions identified and visualized using bar graph

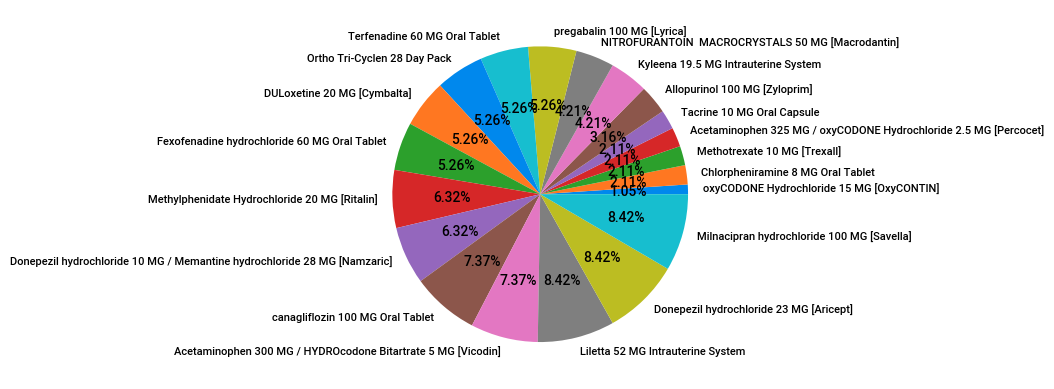
**Top 20 drugs**



These are the top 20 drugs identified and visualized using matplotlib

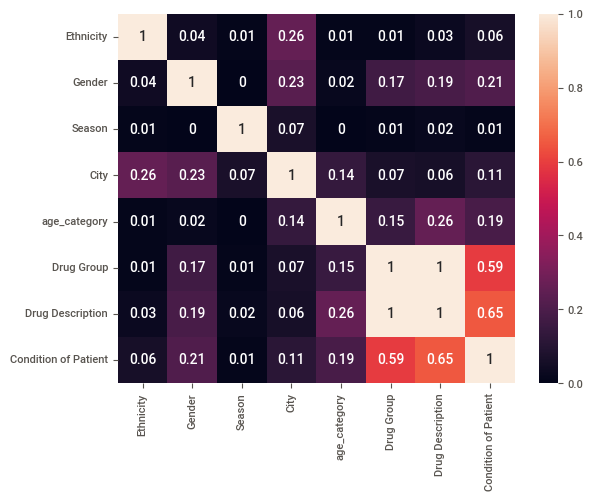
In this Pencillin V potassium is majorly used

**Least 20 drugs**

****

* These are the least 20 drugs identified and visualized using matplotlib

HEAT MAP



* In this heat map we found the corelation between every categorical data columns using Cramer's V function ranges from 0 to 1
* There is a moderate corelation between drug description and the condition of patient and There is a small corelation between age\_category and the condition of the patient
* The other relation between age\_category and drug description