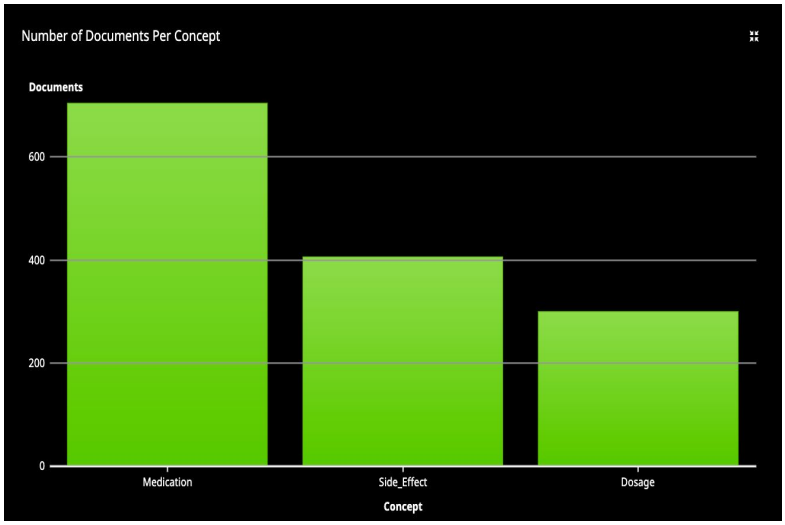
Analytics in SAS Model Studio

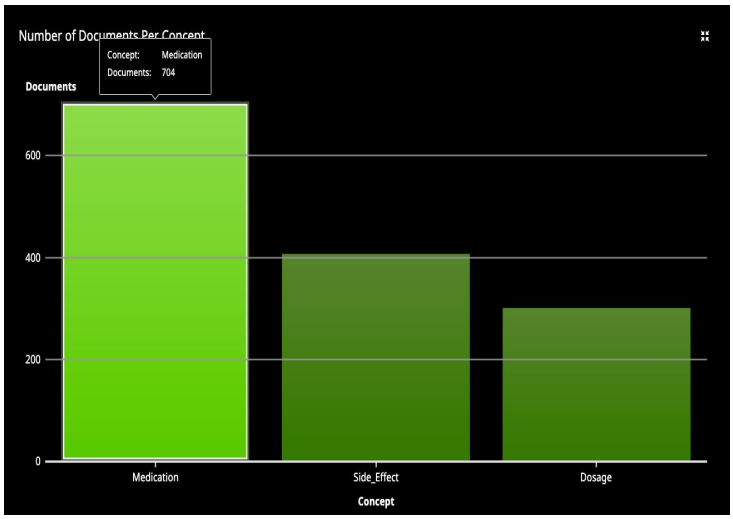
Creator : Yash Masand

1) Charts:

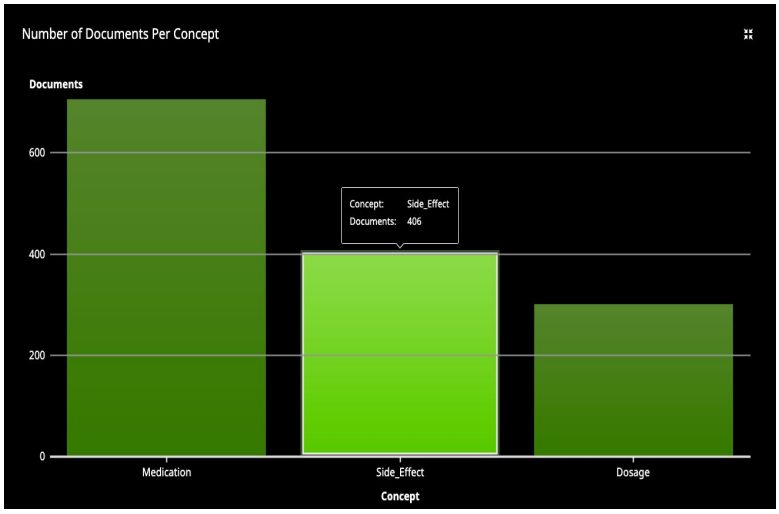
1) Overall Bar Graph



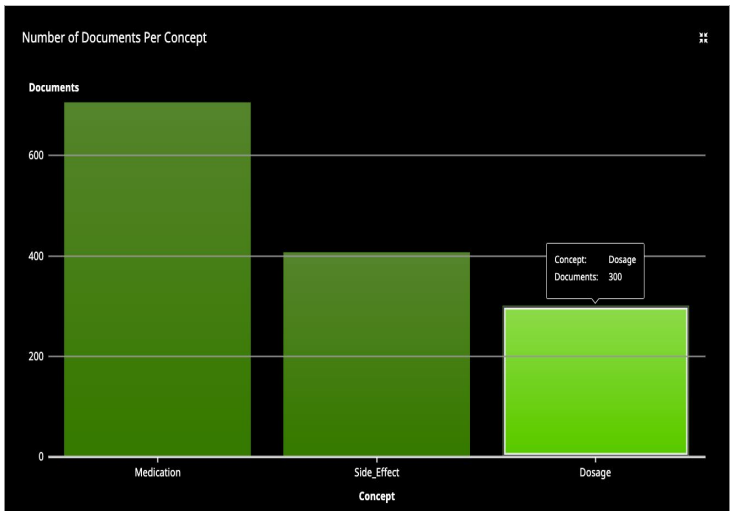
2) 704 Documents of Medication



3) 406 Documents of Side Effects



4) 300 Documents of Dosage

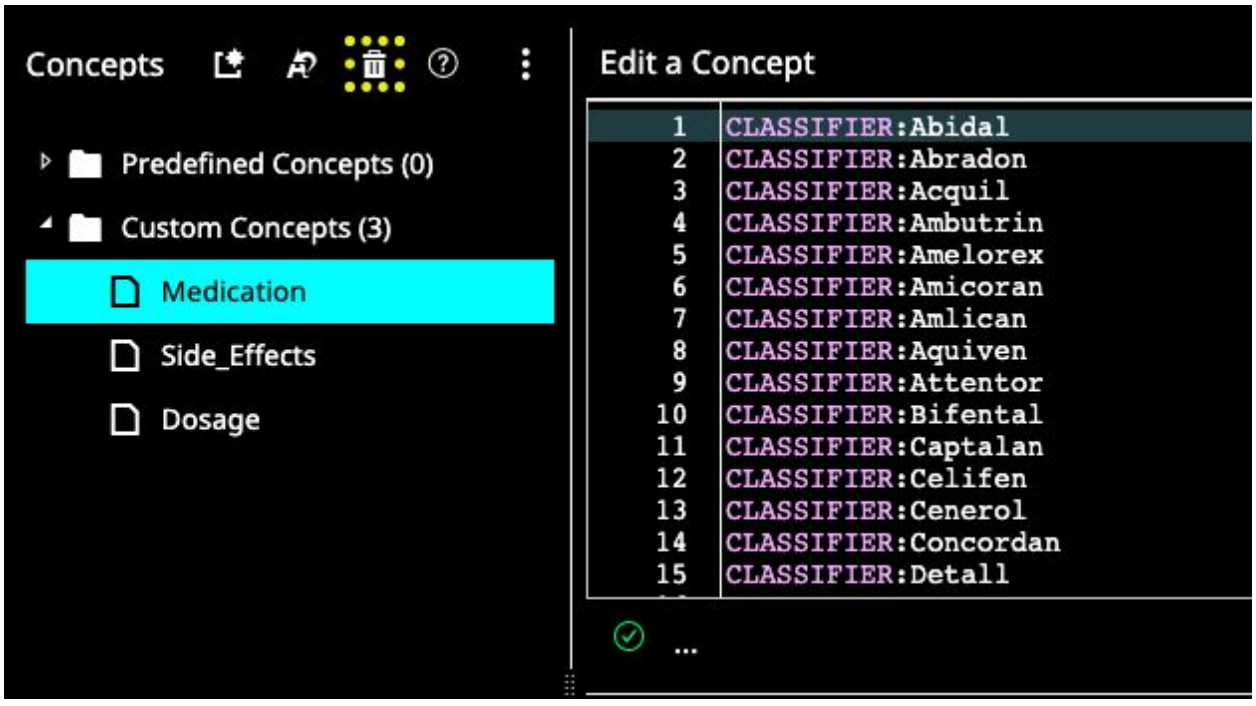


Steps To Produce This Chart:

For this question, we have used SAS Model Studio.

1) After opening the SAS Model Studio, firstly, we click on New Project.
2) We enter the name as “drug_report”, as we’ve done in our case.
3) We choose the Type as “Text Analytics”.
4) We find the Data source from the list, which is “DRUG_REPORTS”.
5) Then, we click on Save.
6) We select DrugReport Variable and choose its role as “Text” on the right-hand side.
7) Next, we go to Pipelines by clicking “Pipelines” from the upper left section.
8) We do a right-click on Concepts and choose to click Open.
9) We make a new Concept under Custom Concepts by clicking “New Concept” button.
10) We add new Concepts, namely - 1) Medication, 2) Side Effects, 3) Dosage.
11) Also, we refer to the text files given in the Readme file for each specific concept and copy all the classifiers of the respective concepts and paste it in the “Edit a Concept” section.
12) Finally, we click on Run Nodes in the upper right section.
13) We shall now close this window.
14) Now, we right click on Concepts again from the Pipeline display and then, choose Results from the drop-down menu.
15) This will ultimately produce three bar graphs & we choose the one we have produced above.

Screenshot of SAS Model Studio Work:

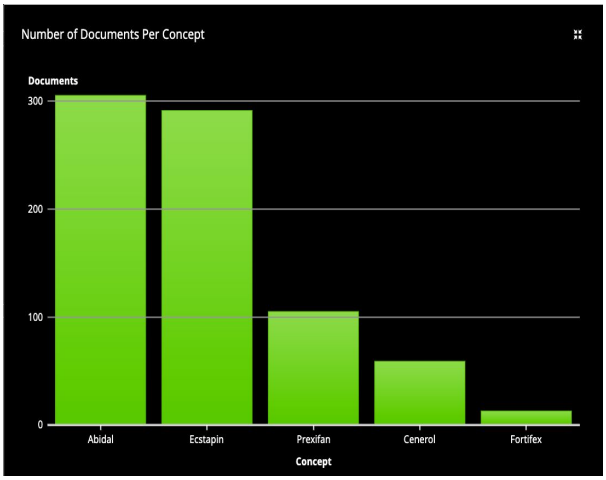


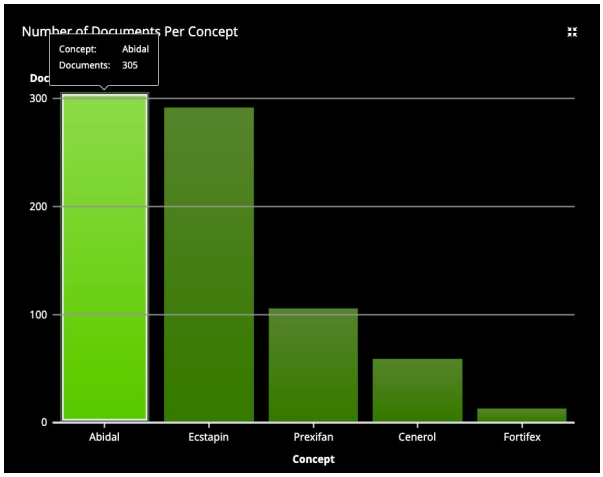
Discussion:

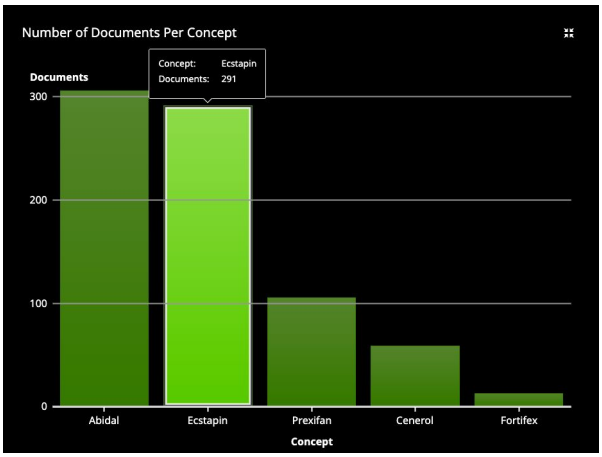
When we hover over Medication on the bar graph of Number of Documents per Concept, there are **704 documents** that contain a **Medication**.
When we hover over Side Effect on the bar graph of Number of Documents per Concept, there are **406 documents** that contain a **Side Effect**.
When we hover over Dosage on the bar graph of Number of Documents per Concept, there are **300 documents** that contain a **Dosage**.

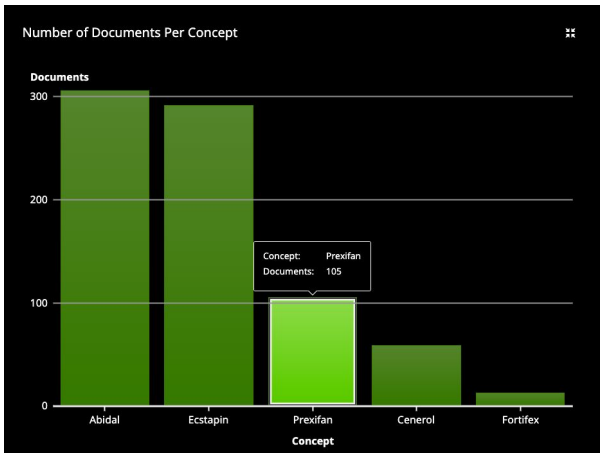
So overall, we observe that Medication is included in the maximum number of documents, followed by Side Effects and lastly, Dosage.

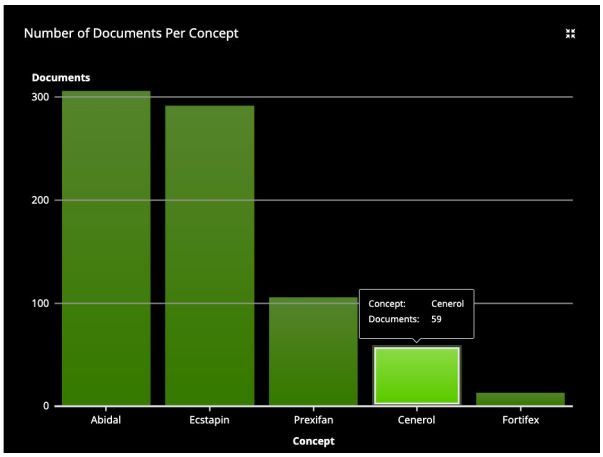
2) Charts:

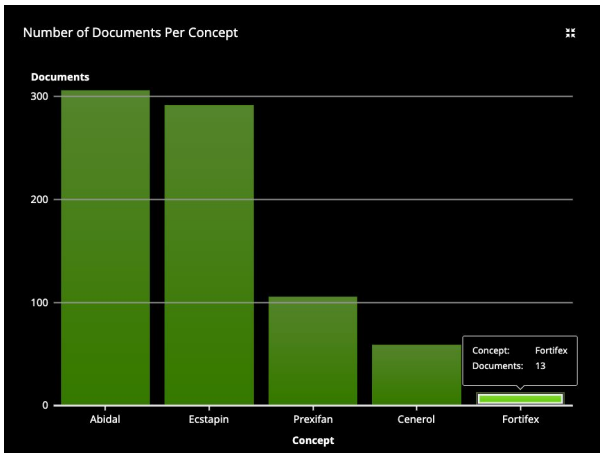

Number of Documents Per Concept
Documents
300
200
100
0
Abidal
Ecstapin
Prexifan
Cenerol
Fortifex
Concept


Number of Documents Per Concept
Concept: Abidal
Documents: 305
300
200
100
0
Abidal
Ecstapin
Prexifan
Cenerol
Fortifex
Concept


Number of Documents Per Concept
Concept: Ecstapin
Documents: 291
Documents
300
200
100
0
Abidal
Ecstapin
Prexifan
Cenerol
Fortifex
Concept


Number of Documents Per Concept
Documents
300
200
Concept: Prexifan
Documents: 105
100
0
Abidal
Ecstapin
Prexifan
Cenerol
Fortifex
Concept


Number of Documents Per Concept
Documents
300
200
Concept: Cenerol
Documents: 59
100
0
Abidal
Ecstapin
Prexifan
Cenerol
Fortifex
Concept

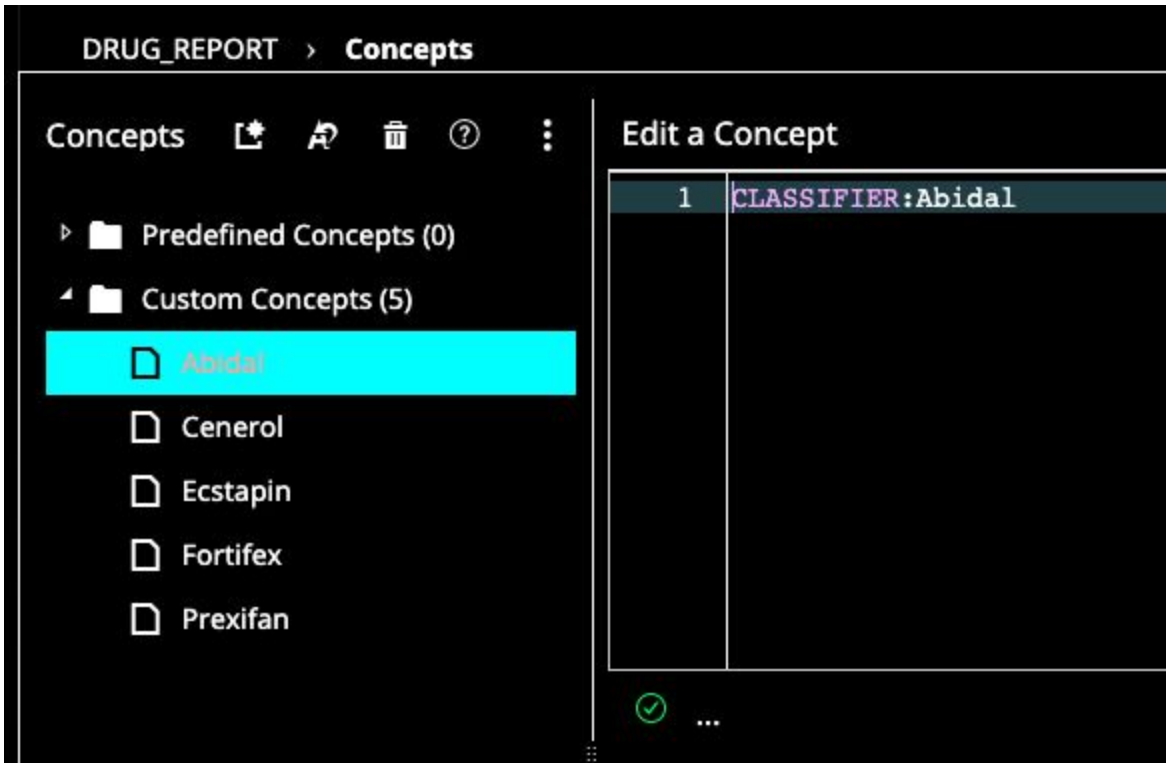

Number of Documents Per Concept
Documents
300
200
100
Concept: Fortifex
Documents: 13
0
Abidal
Ecstapin
Prexifan
Cenerol
Fortifex
Concept

Steps To Produce This Chart:

For this question, we have used SAS Model Studio.

1) In SAS Model Studio,firstly we right-click on Concepts from the Pipeline display and choose Open.
2) We then make new Concepts under Custom Concepts by clicking “New Concept” button.
3) We add new Concepts, namely - 1) Abidal, 2) Cenerol, 3) Ecstapin 4) Fortifex 5) Prexifan.
4) We are supposed to input ‘Classifier:Medication Name’ for each specific medication in the ‘Edit a Concept’ Tab.
5) We then, click on Run Node from the upper right section to execute this work.
6) Click on “Close” from the upper right section, beside “Run Node”, after execution.
7) Now, we right-click on Concepts again from the Pipeline display & click on Results from the drop-down menu.
8) This will produce three bar graphs. We choose the one we have produced above.

Screenshot of SAS Model Studio Work:



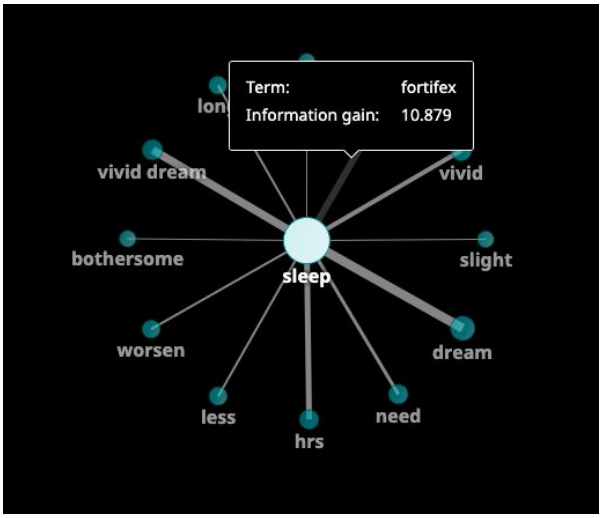
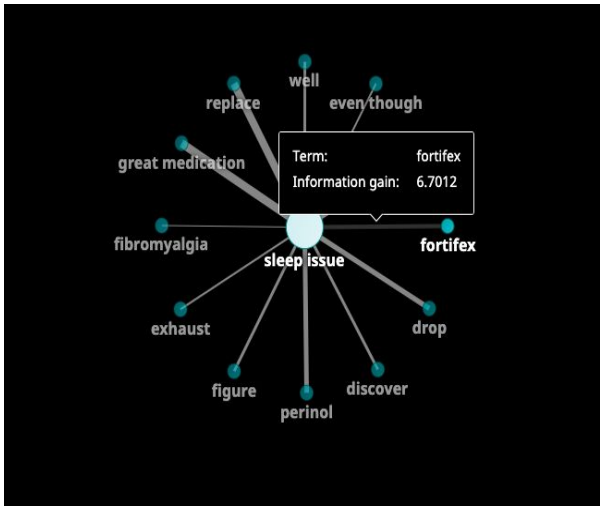
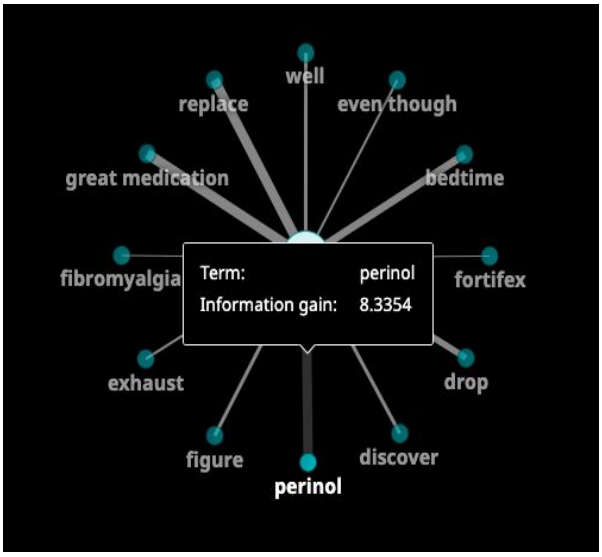
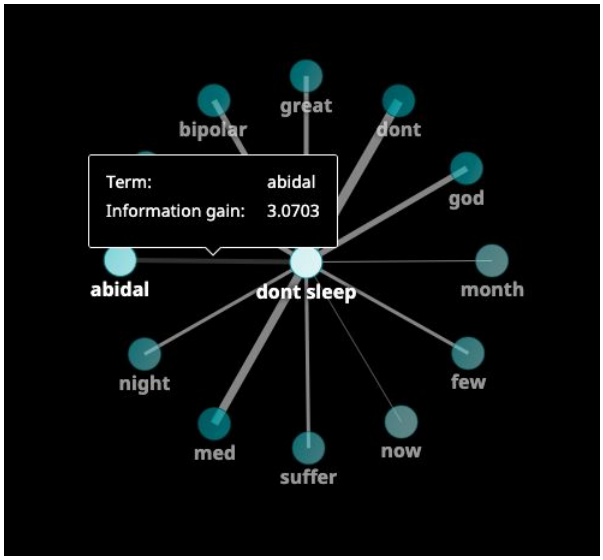
Discussion:

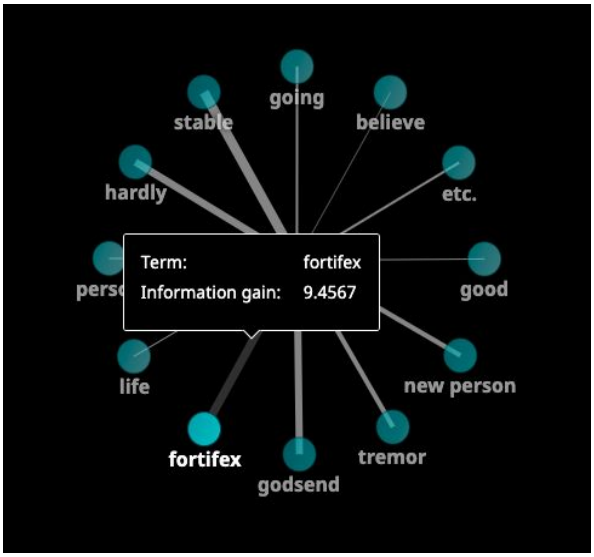
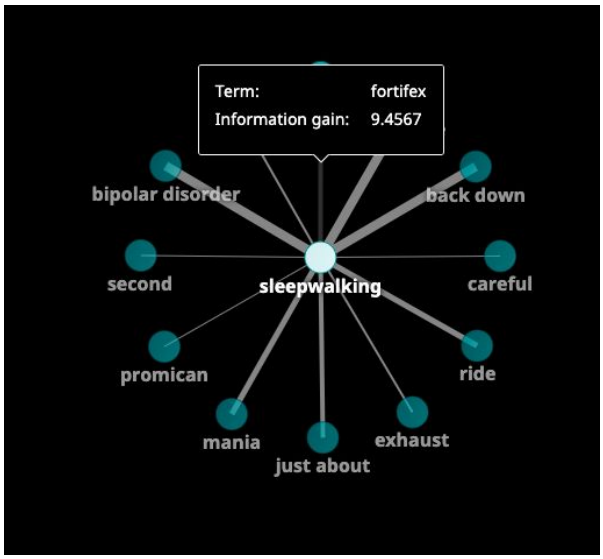
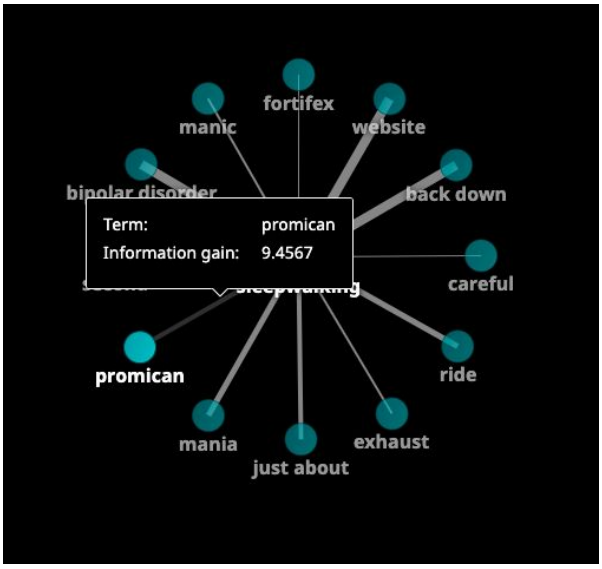
From the above Bar Graphs, we can compare and observe that -

1) **Abidal** is mentioned in a total of **305 documents.**
2) **Cenerol** is mentioned in a total of **59 documents.**
3) **Ecstapin** is mentioned in a total of **291 documents**.
4) **Fortifex** is mentioned in a total of **13 documents.**
5) **Prexifan** is mentioned in a total of **105 documents.**

Thus, after comparing the counts for each of the five medications, Abidal is highest mentioned medication out of all the five medications whereas, Fortifex is the least mentioned medication out of all the five medications.

3) Chart:

1) sleep (fortifex)	2) sleep issue (fortifex)
	
3) sleep issue (perinol)	**4) dont sleep (abidal)**
	

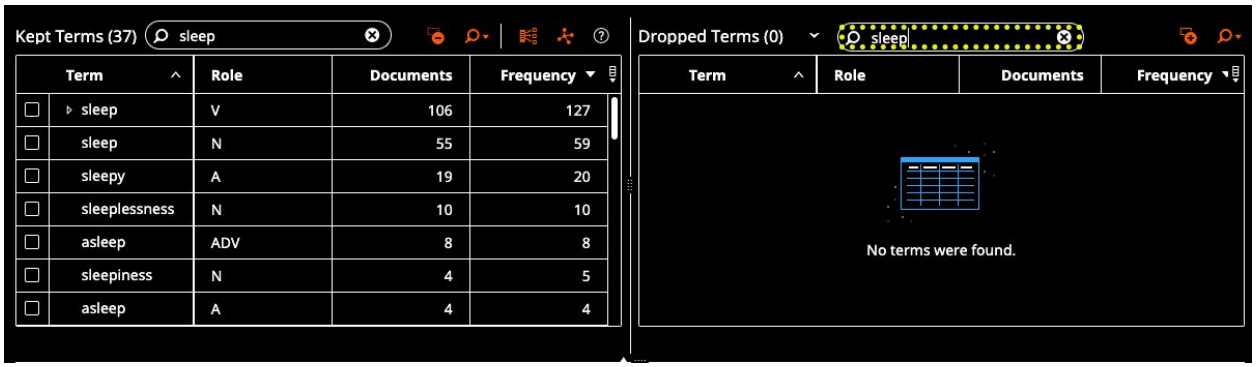
5) sleepwalk (fortifex)	6) sleepwalking (fortifex)
	
7) sleepwalking (promican)	
	

Steps To Produce This Chart:

For this question, we have used SAS Model Studio.

1) In SAS Model Studio, firstly, we right-click on Text Parsing from the Pipeline display and choose “Open”.
2) We are displayed with 2 sections - Kept Terms (on the left) & Dropped Terms (on the right).
3) We make a search query in Kept Terms Tab using keyword - “Sleep”. (since we are interested to find out about sleep)
4) We open each respective Term Map by clicking the “Show term map” icon and discover the name of Medications associated with sleep in each Term Map.
5) In Dropped Terms, we again select all necessary Terms that indicate sleep issues and add them to Kept Terms by clicking “Keep terms” icon.
6) Finally, we check the new Term Maps as well for the names of Medications associated with sleep.

Screenshot of SAS Model Studio Work:



Discussion:

As per the Term Maps that we have produced, we have noticed that Medications associated with sleep issues are: Fortifex, Perinol, Abidal and Promican.

Their respective Information gain values are as follows:

1) **Fortifex** = 9.4567 , 9.4567, 10.879 , 6.7012 (numerous values due to multiple roles)
2) **Perinol** = 8.3354
3) **Abidal** = 3.0703
4) **Promican** = 9.4567

Information Gain tells us how much "information" a feature explains about a particular class. Here, the classes have been mentioned in the table along with their nodes. So we can say that,

For class **sleep, fortifex** has an information gain of 10.879. (Role = Noun)
For class **sleep issue, fortifex** has an information gain of 6.7012. (Role = nIpNounGroup)
For class **sleep issue, perinol** has an information gain of 8.3354.
For class **dont sleep, abidal** has an information gain of 3.0703.
For class **sleepwalk, fortifex** has an information gain of 9.4567. (Role = Verb)
For class **sleepwalking, fortifex** has an information gain of 9.4567. (Role = Noun)
For class **sleepwalking, promican** has an information gain of 9.4567.

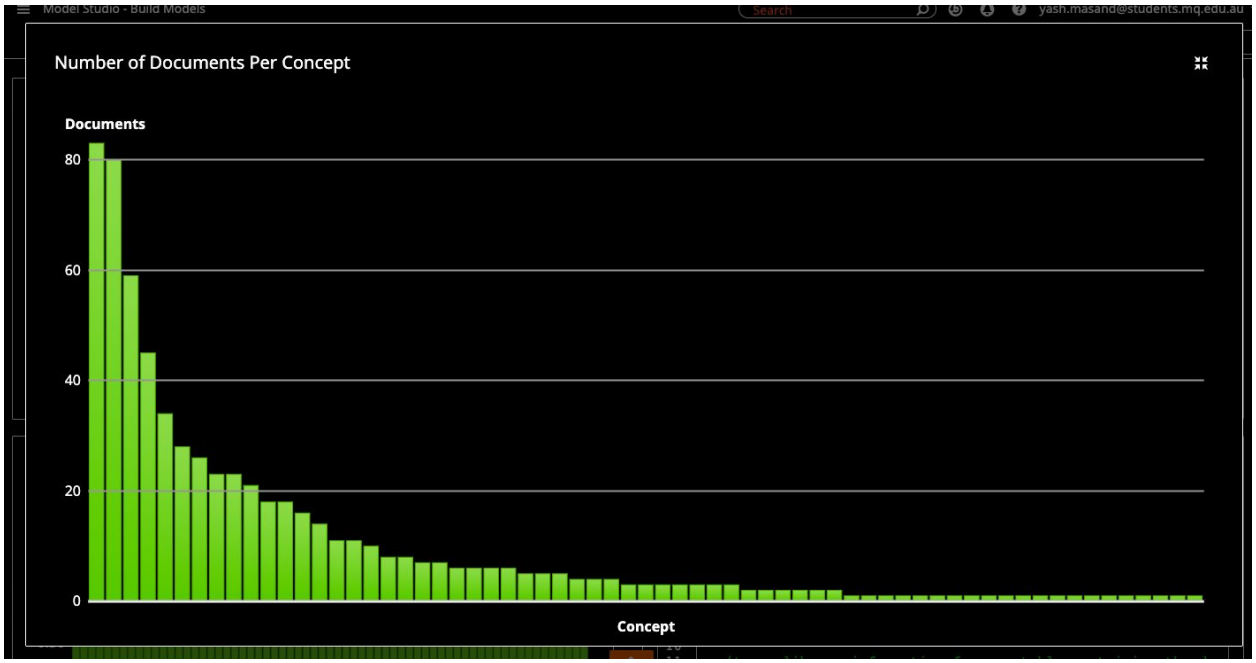
Here, the information gain works as the degree of association which helps us in comparing each degree of association of the respective medications.

Also, we observe multiple information gain values for fortifex due to differing roles of the word "sleep" in the posts of the customers.

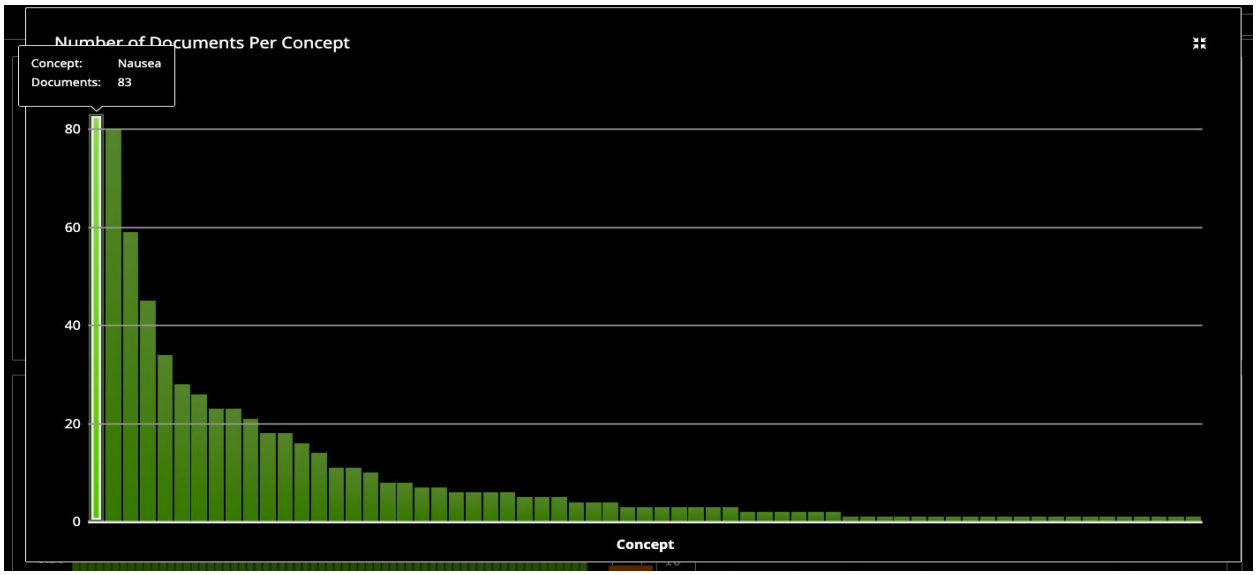
Overall, we can conclude that out of all medications, the medication "Fortifex" is linked with maximum sleep issues amongst the patients followed by "Promican", then "Perinol" and lastly, "Abidal".

4) Charts:

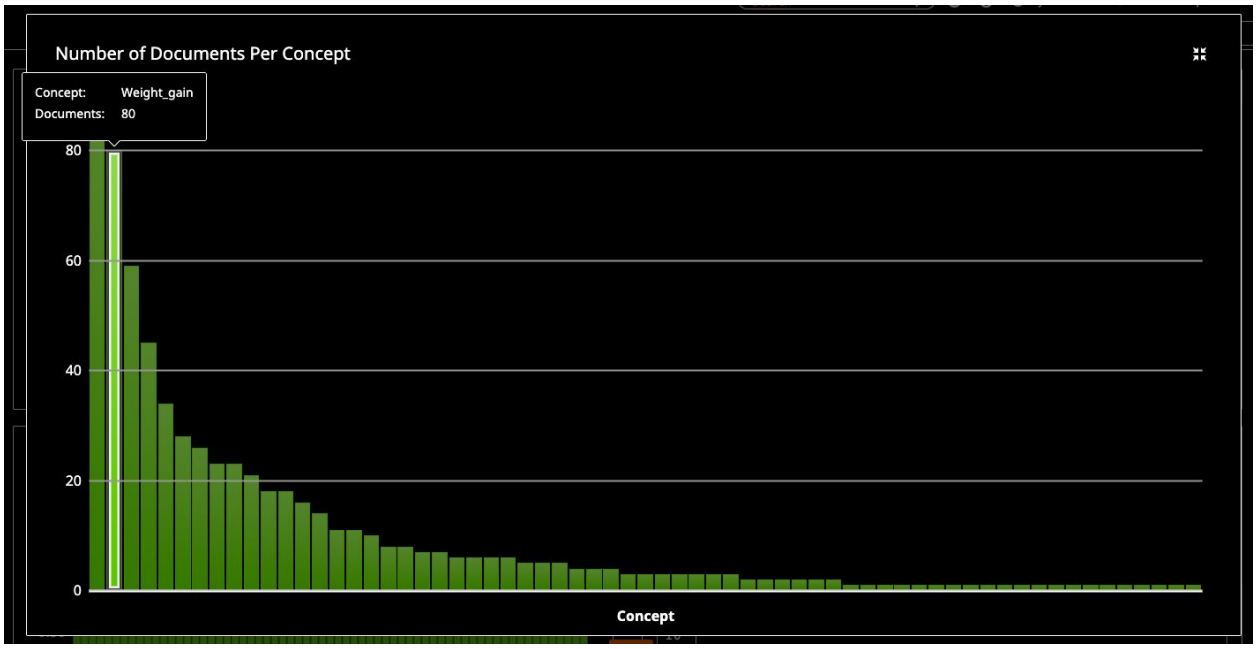
Overall Graph:



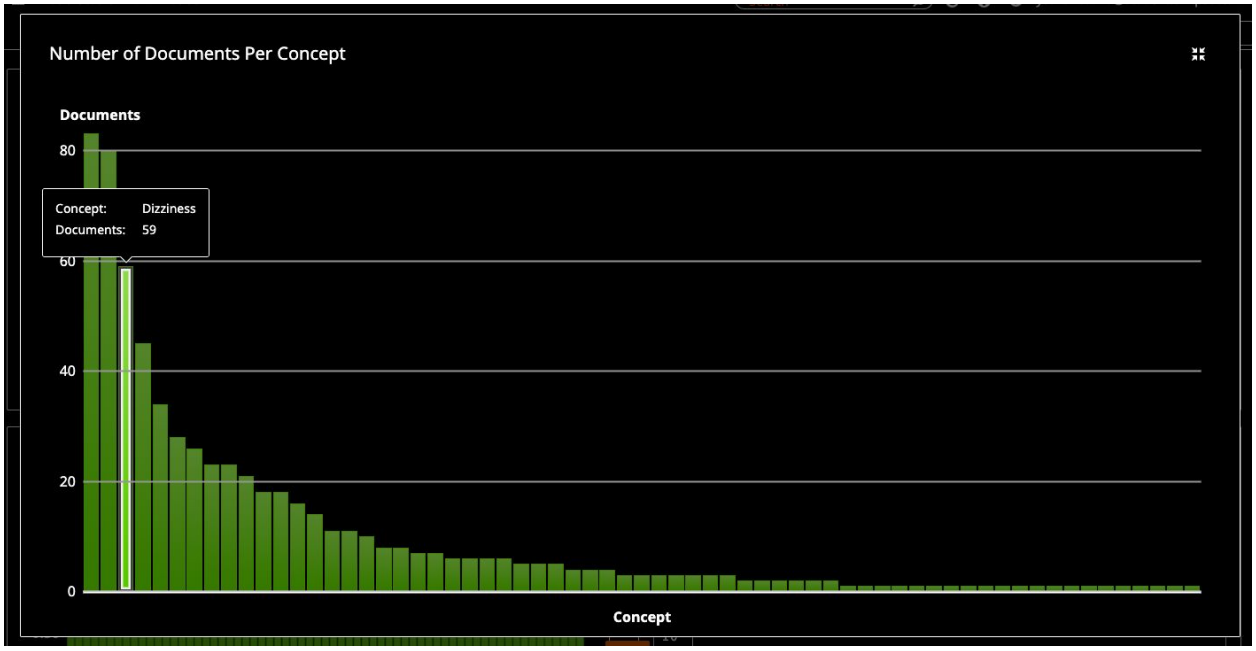
Nausea:



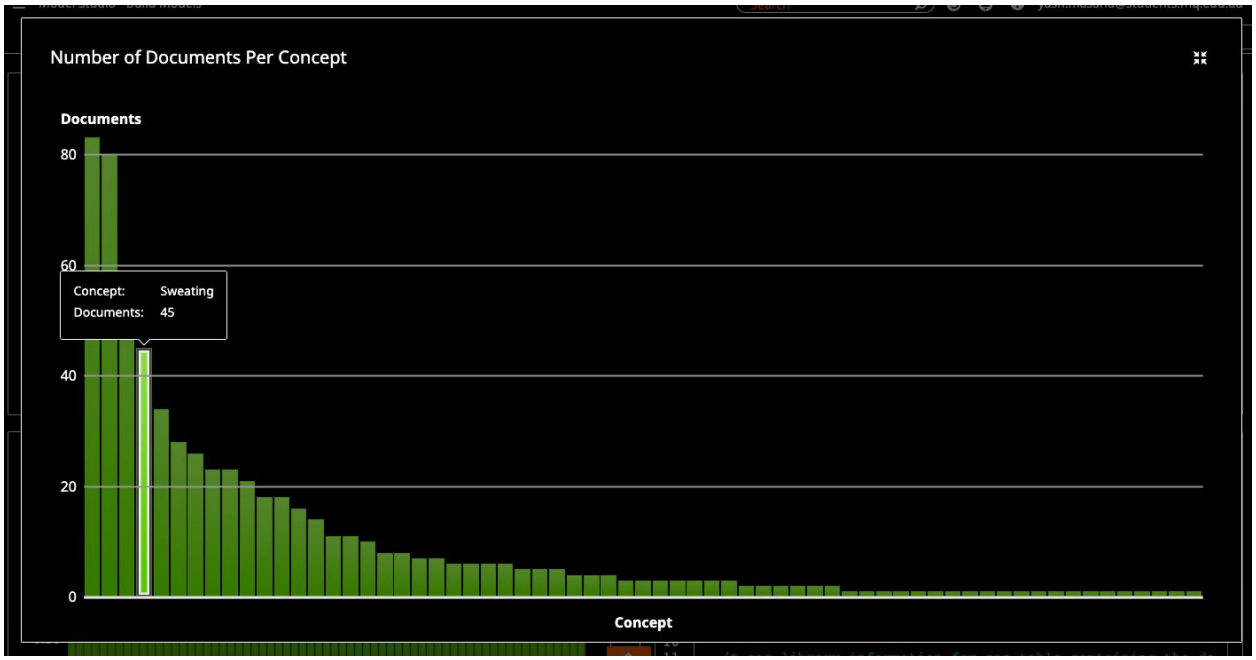
Weight Gain:



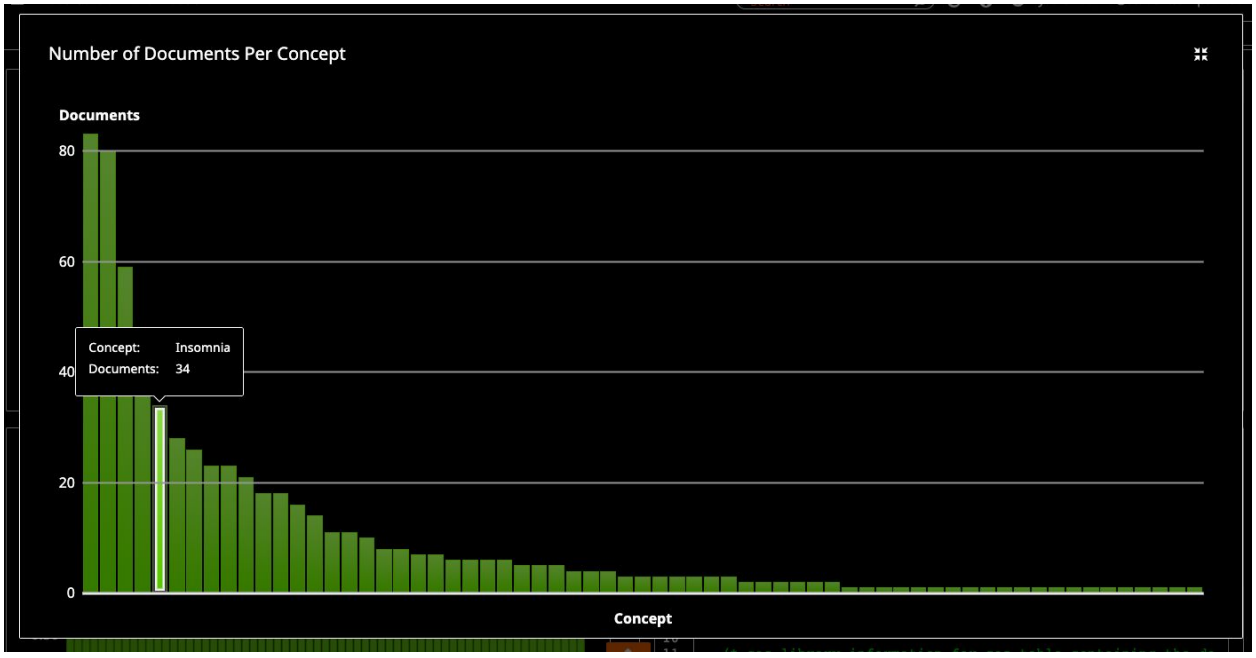
Dizziness:



Sweating:



Insomnia:

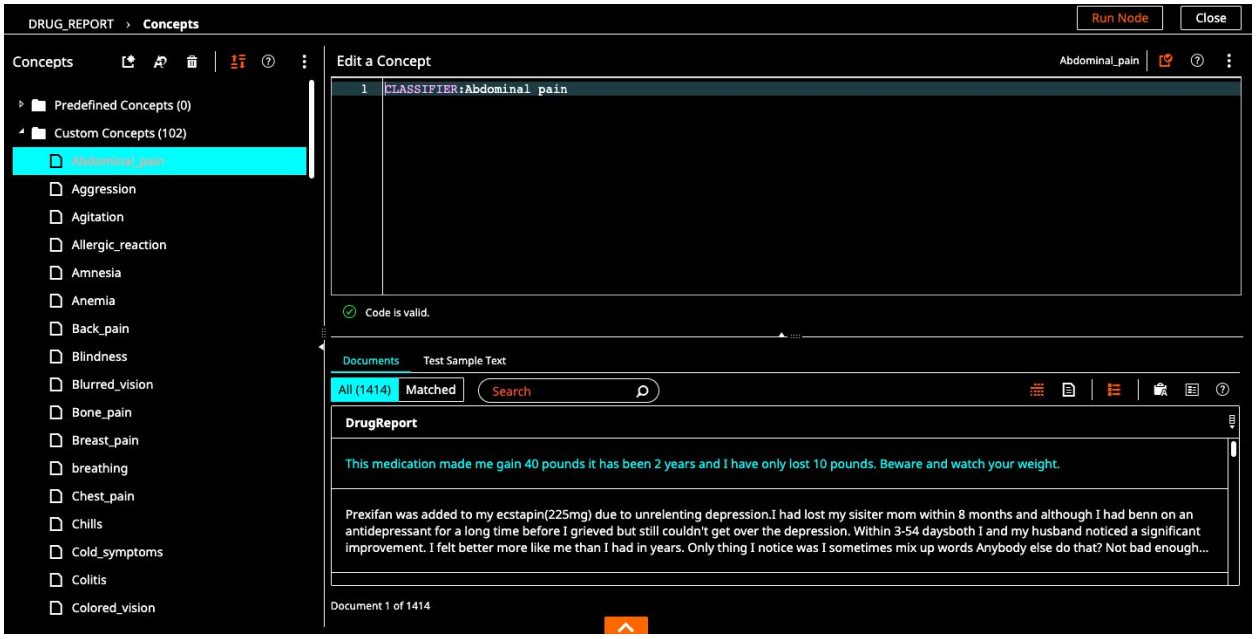


Steps To Produce This Chart:

For this question, we have used SAS Model Studio.

1) In SAS Model Studio, we first right-click Concepts from the Pipeline display and choose Open from the drop-down menu.
2) Next, we make new Concepts under Custom Concepts by clicking “New Concept” button.
3) Then, we add new Concepts, namely all the 102 Side Effects included in “SideEffectsConcept.txt” file given in the Readme file.
4) After that, we copy all the classifiers of the respective concepts and paste it in them under the newly created Concepts in the “Edit a Concept” tab.
5) After putting all the classifiers correctly as shown in the screenshot below, we click on “Run Node” command in the upper right section.
6) Now, we close the window by choosing “Close” in the upper right section as well.
7) On the Pipeline display, we again right-click on Concepts again and click on “Results” from the drop-down menu.
8) This will produce three bar graphs and we choose the one we have produced above to interpret our results, given above.

Screenshot of Software Work:



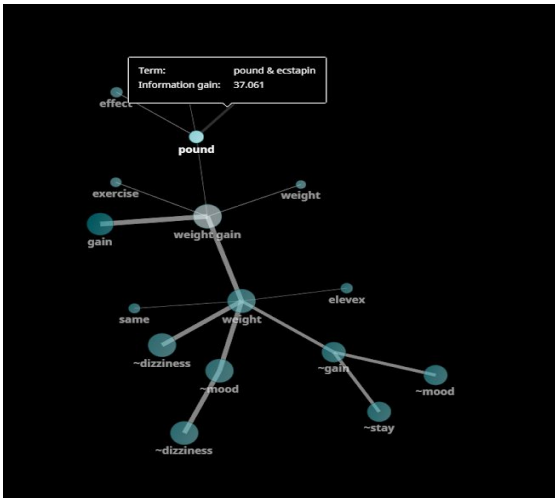
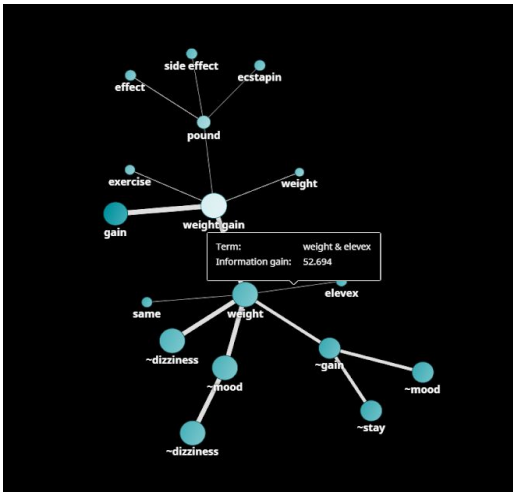
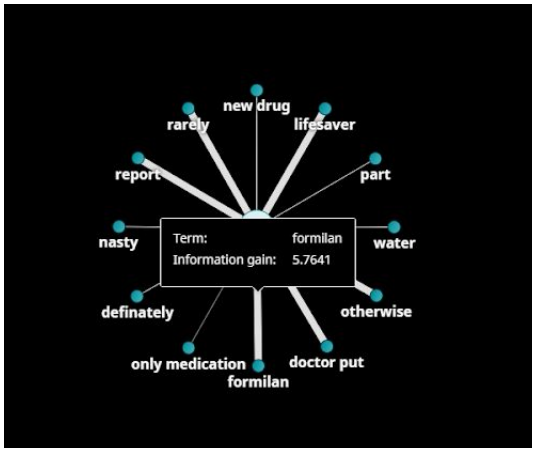
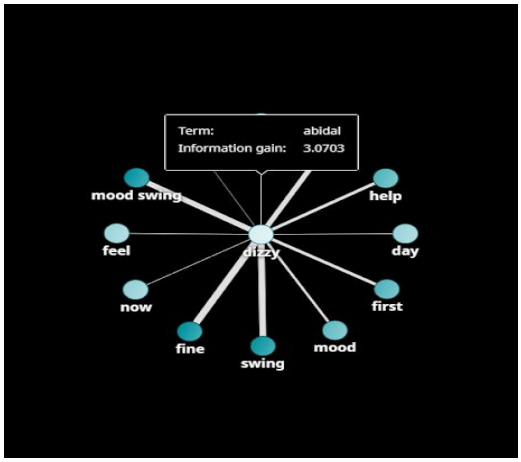
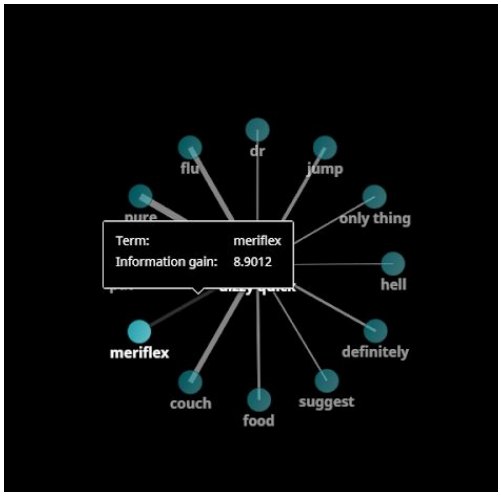
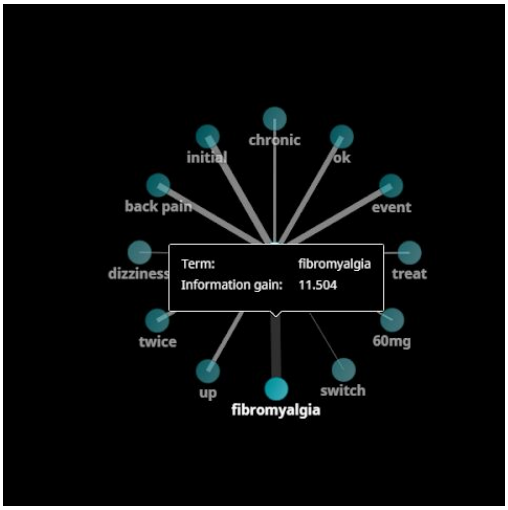
Discussion:

The main concerns expressed by most patients in all the posts are:

1) **Nausea** - (Contained in **83 Documents**)
2) **Weight Gain** - (Contained in **80 Documents**)
3) **Dizziness** - (Contained in **59 Documents**)
4) **Sweating** - (Contained in **45 Documents**)
5) **Insomnia** - (Contained in **34 Documents**)

This can be observed with the help of above bar graphs which includes the counts of the Number of Documents per Concept (in this case, Side Effects). In order to read the bar graph counts, we just hover over the first 5 bars to see which "Side Effect" they represent along with the number of counts in the total documents in order to see the Top 5 Concerns of the customers according to their posts.

5) Charts:

1) weightgain - ecstapin	2) weight - elevex
	
3) weight - formilan	**4) dizzy - abidal**
	
5) dizzy quick - meriflex	**6) initial dizziness - fibromyalgia**
	

Steps To Produce This Chart:

For this question, we have used SAS Model Studio.

1) In SAS Model Studio, we right-click on Text Parsing from the Pipeline display and Click "Open" from the drop-down menu.
2) A new window opens up which consists of "Kept Terms" (on the left) and "Dropped Terms" (on the right). We search in Kept Terms Tab using keyword - "**Nausea**" as we are looking for medications associated with Nausea.
3) We open each respective Term Map by clicking on the "Show term map" icon and discover the name of Medications associated with Nausea.
4) Next, in Dropped Terms, we select all necessary Terms that indicate nauseatic issues by again passing a search query.
5) We click on "Run Node" from the upper right section to execute this action.
6) Lastly, we check the new Term Maps as well for the associations.

7) After choosing the necessary term maps related to Nausea, we close this window and return back to "Text Parsing - Manage Terms" Window.
8) We again, search in Kept Terms Tab using keyword - "**Weight Gain**".
9) Then, we open each respective Term Map and discover the name of Medications associated with Weight Gain. (same as Nausea Process)
10) In Dropped Terms, Select all necessary Terms that indicate weight gain issues.
11) We again click on "Run Node" from the upper right section to execute the action.
12) Finally, we check the new Term Maps as well.

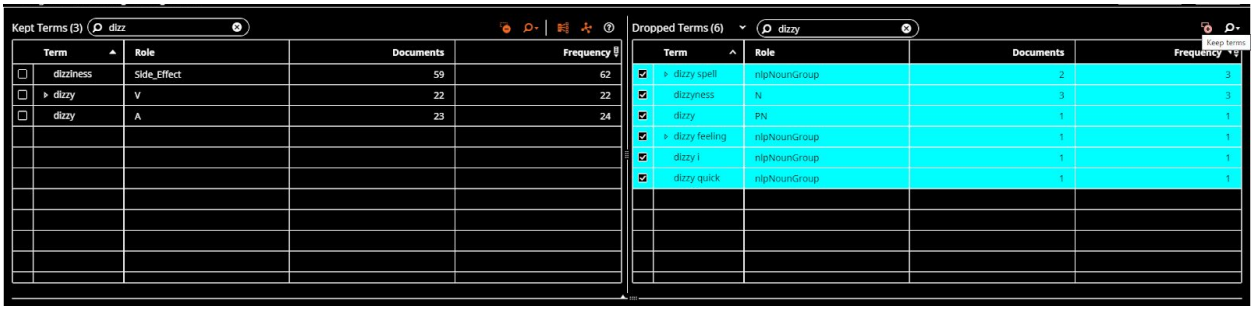
13) This process is repeated again for Dizziness again.
14) After choosing the necessary term maps for Weight Gain, we close the window again.
15) Next, we search in Kept Terms Tab using keyword - "**Dizziness**".
16) Again, we open each respective Term Map and discover the name of Medications associated with Dizziness.
17) In Dropped Terms, we select all necessary Terms that indicate dizziness issues.
18) We make sure to Run the node.
19) Ultimately, we check the new Term Maps as well.

Screenshot of Software Work:

Nausea:

Weight Gain:

Dizziness:



Discussion:

We have produced Term Maps for the "3 Top Most Common Side Effects" in order to understand what Medications lead to these side effects.

From question 4, we have already discovered that the top most common Side Effect is Nausea, the second most common Side Effect is Weight gain, and the third most common Side Effect is Dizziness.

After performing the Term Map analysis, for **Nausea**, we weren't able to discover much from the Term Maps regarding most common medications.

Secondly, we have discovered that for **Weight Gain**, the associated medications are **Ecstapin, Elevex and Formilan** where Elevex is the first, Ecstapin is second and Formilan is the last in terms of degree of association (information gain).

Lastly for **Dizziness**, the associated medications are **Fibromyalgia, Meriflex and Abidal**, where Fibromyalgia is the first, Meriflex is the second and Abidal is the last in terms of information gain.