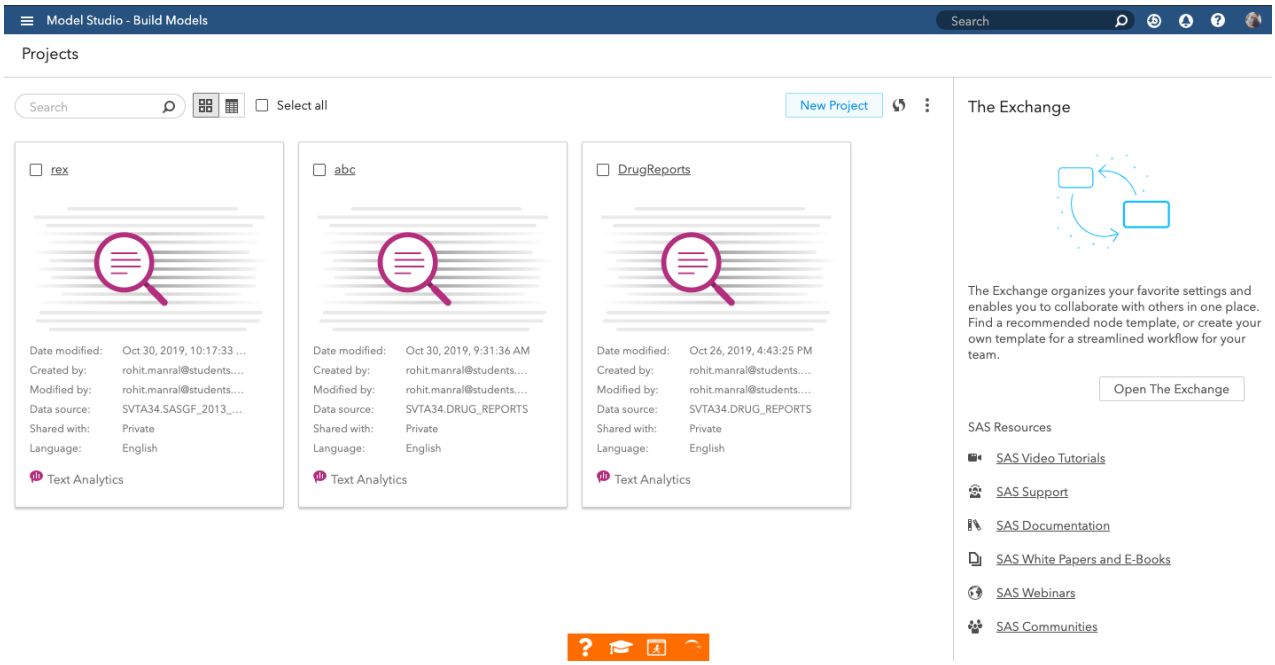
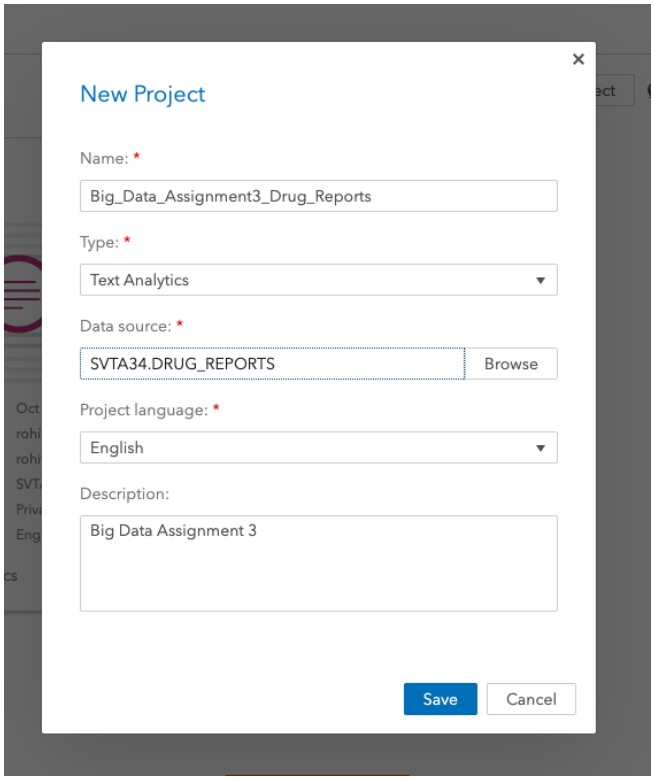
Assignment 3 - Data Analysis

Using SAS Viya Learners

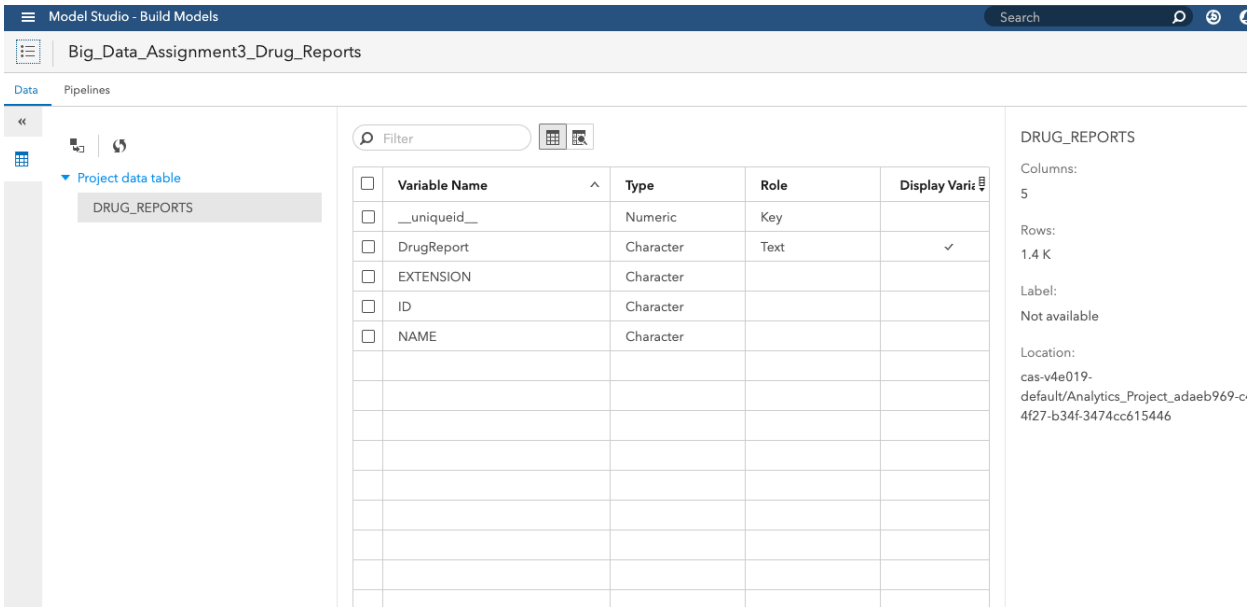
Step 1: Just after signing up with SAS, you will a screen like this:



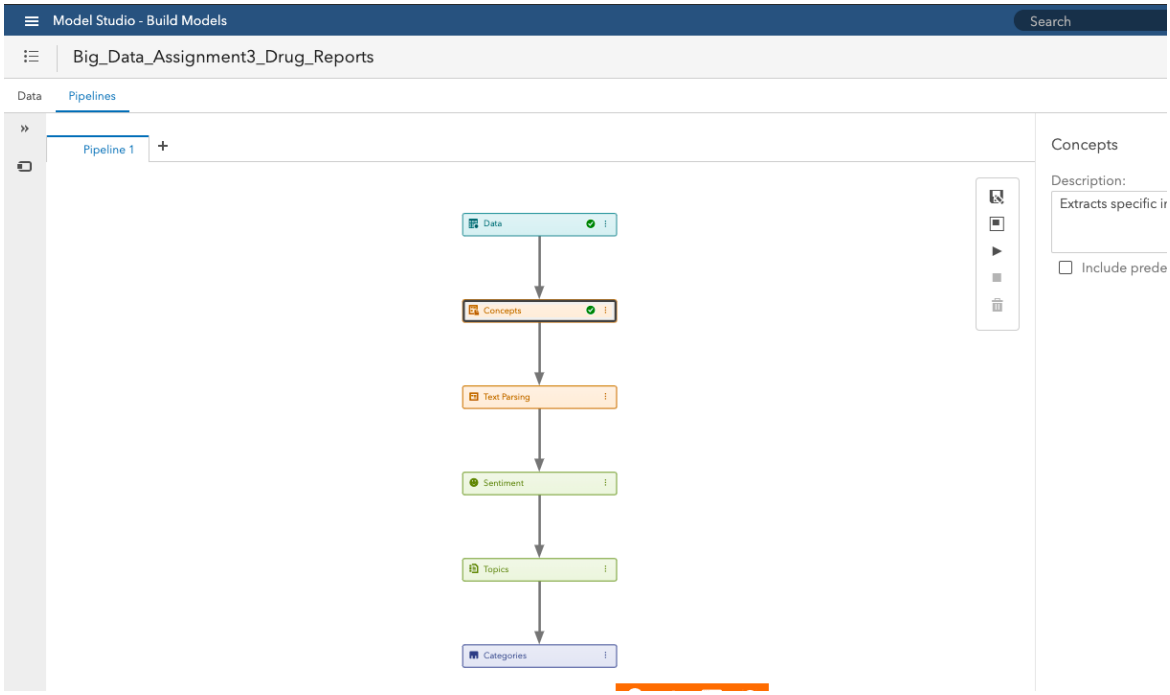
Step 2: Select *New → Model Studio Project* or *New Project*
Then, a new sub-screen will be popped up where you have to fill the necessary details of our New Project.



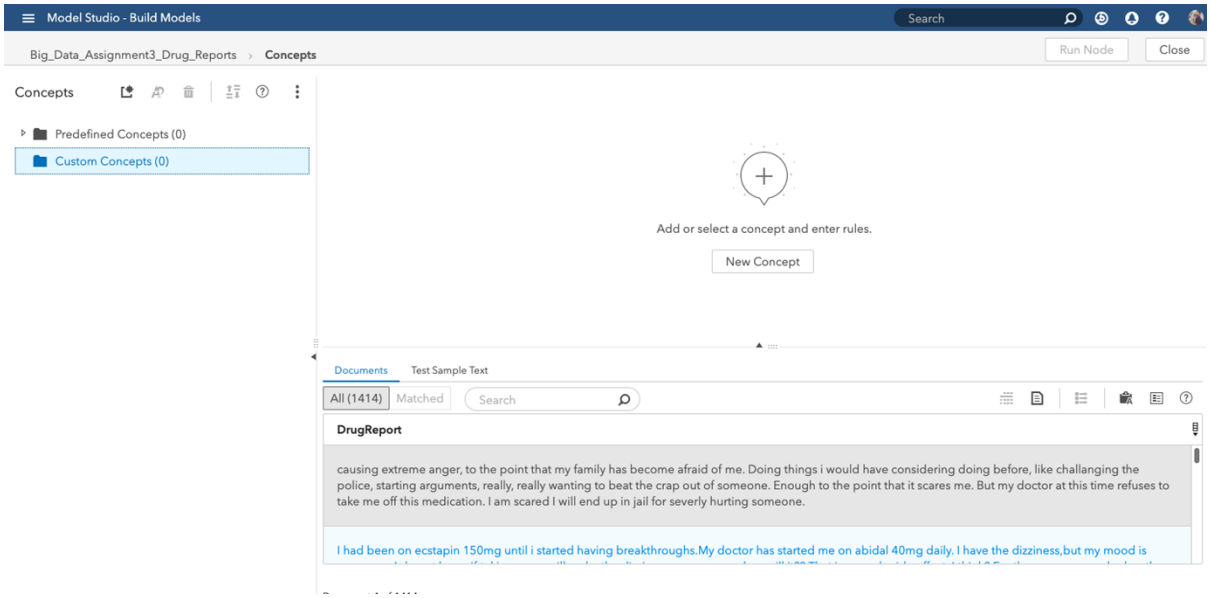
Then, *Text Analysis* will be selected as Type and *SVTA34.DRUG_REPORTS* as a Data Source. After that, I selected *DrugReport* variable as a *Text*.



I went to Pipelines from Data and ran Data & Concepts.



Then, I opened Concepts.



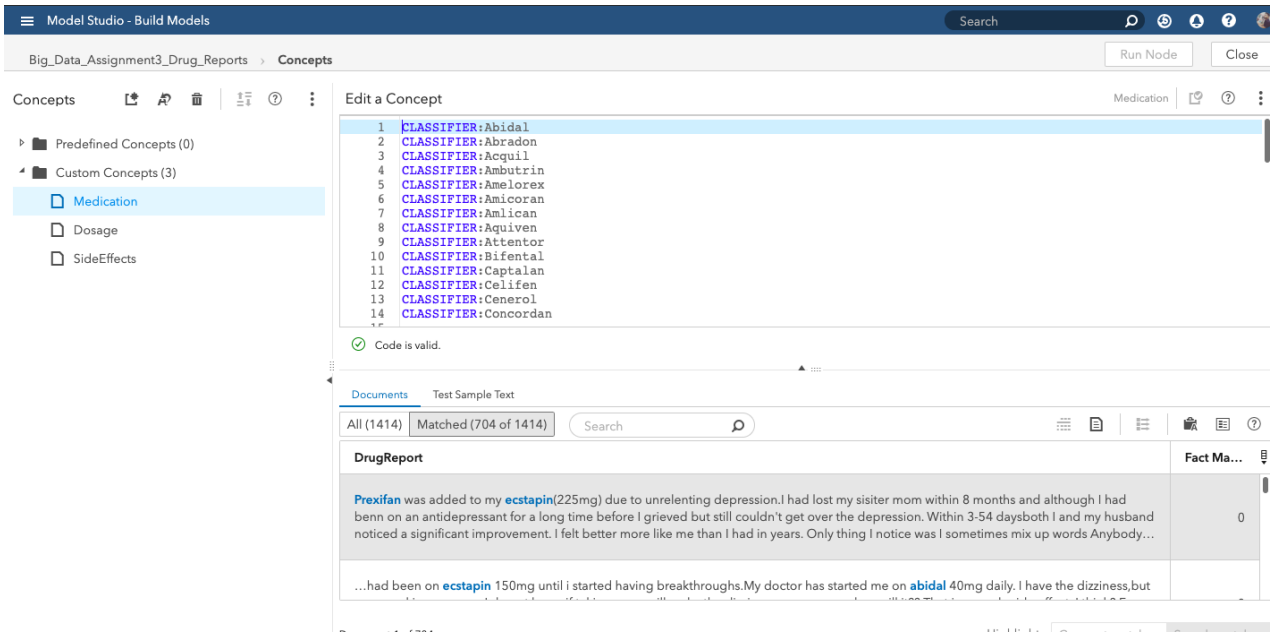
Finally, here we will start completing the 5 given tasks.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

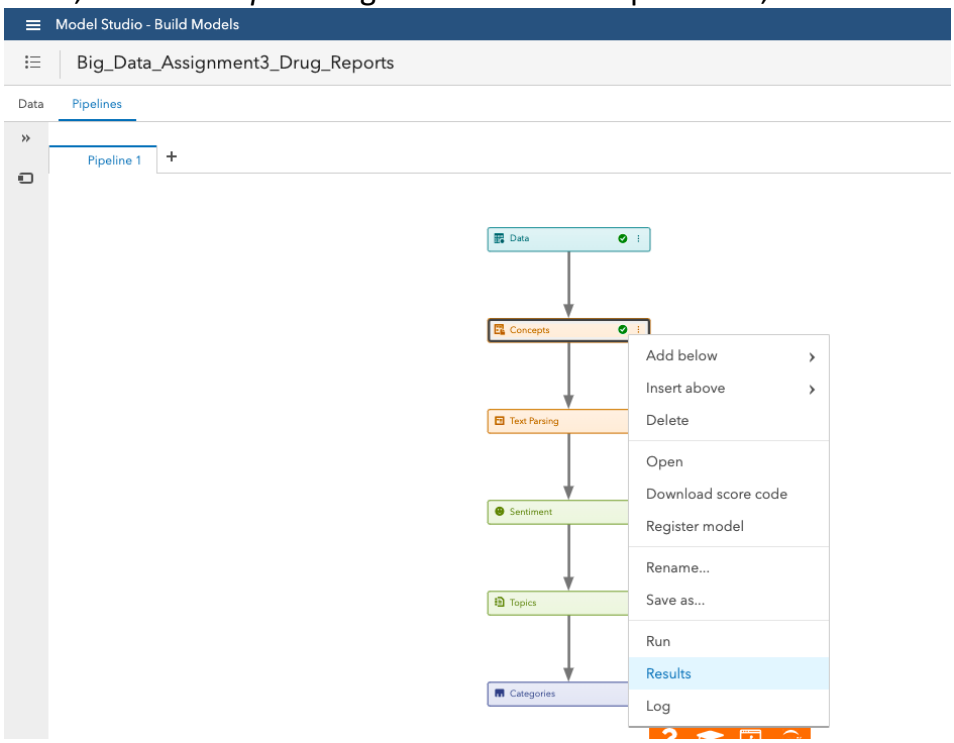
Task 1 - How many documents mention a medication, or a dosage, or a side effect?

For Task 1, I used only SAS Viya for Learners.

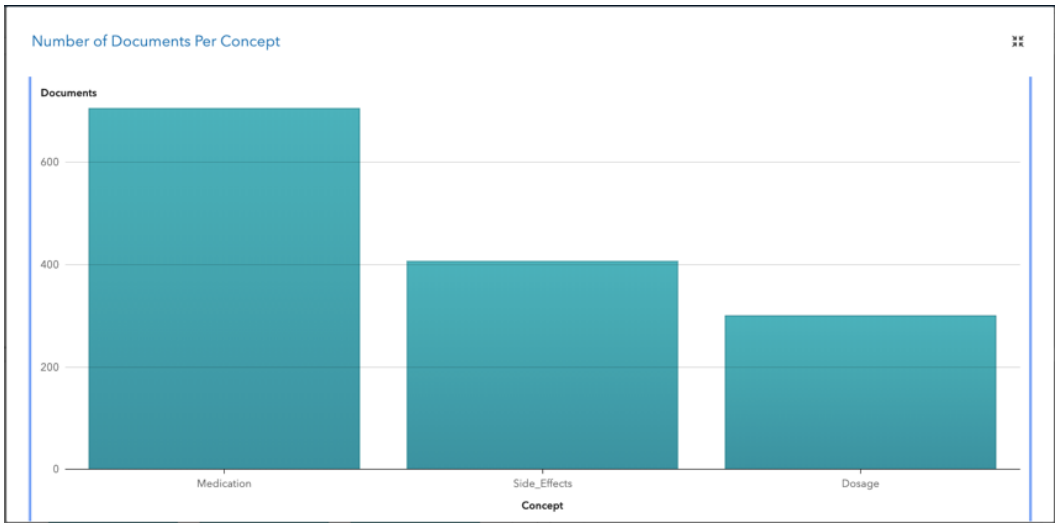
So, I created 3 different concepts (medication, dosage, & side effect) and pasted the required classifiers in all the 3 concepts from the three given text files. In addition, I validated the rules and ran node for every concept.



Then, I went to *Pipeline* again and *ran* Concepts. Then, selected *Results* for Concepts.



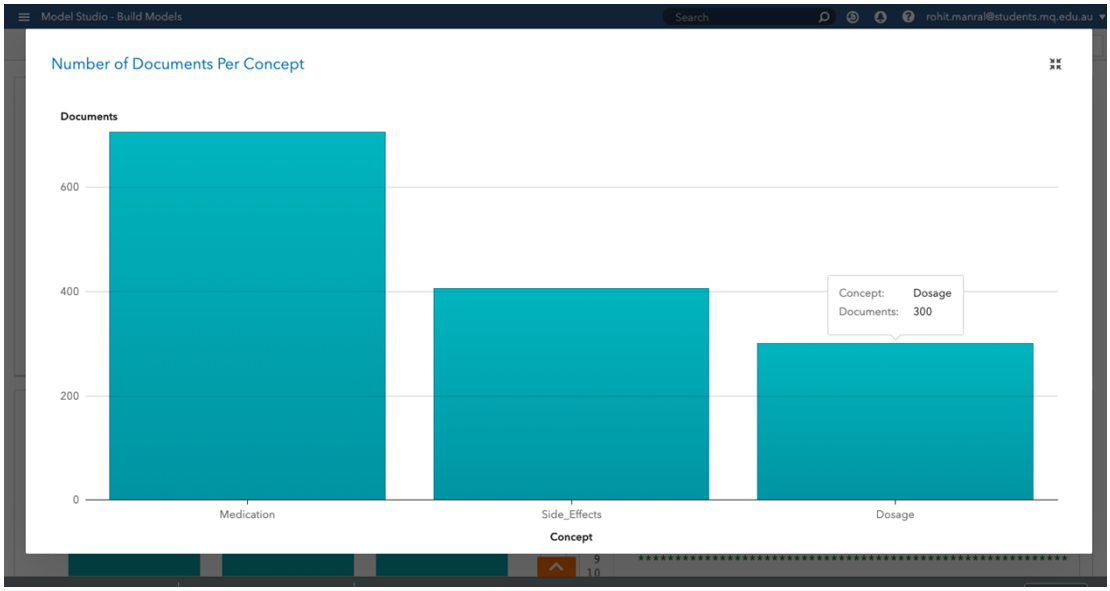
Finally, we will get the bar graph representing the count for Medication, Dosage, & SideEffects.



By scrolling or hovering over the different bars we can get the count of each element i.e. Medication, Dosage, & SideEffects.

- **Scrolling or hovering over Medication Bar**

- **Scrolling or hovering over Dosage Bar**



- **Scrolling or hovering over SideEffects Bar**

Therefore, ***704*** documents mention ***Medication***, ***300*** documents mention ***Dosage***, & ***406*** documents mention ***SideEffects***.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

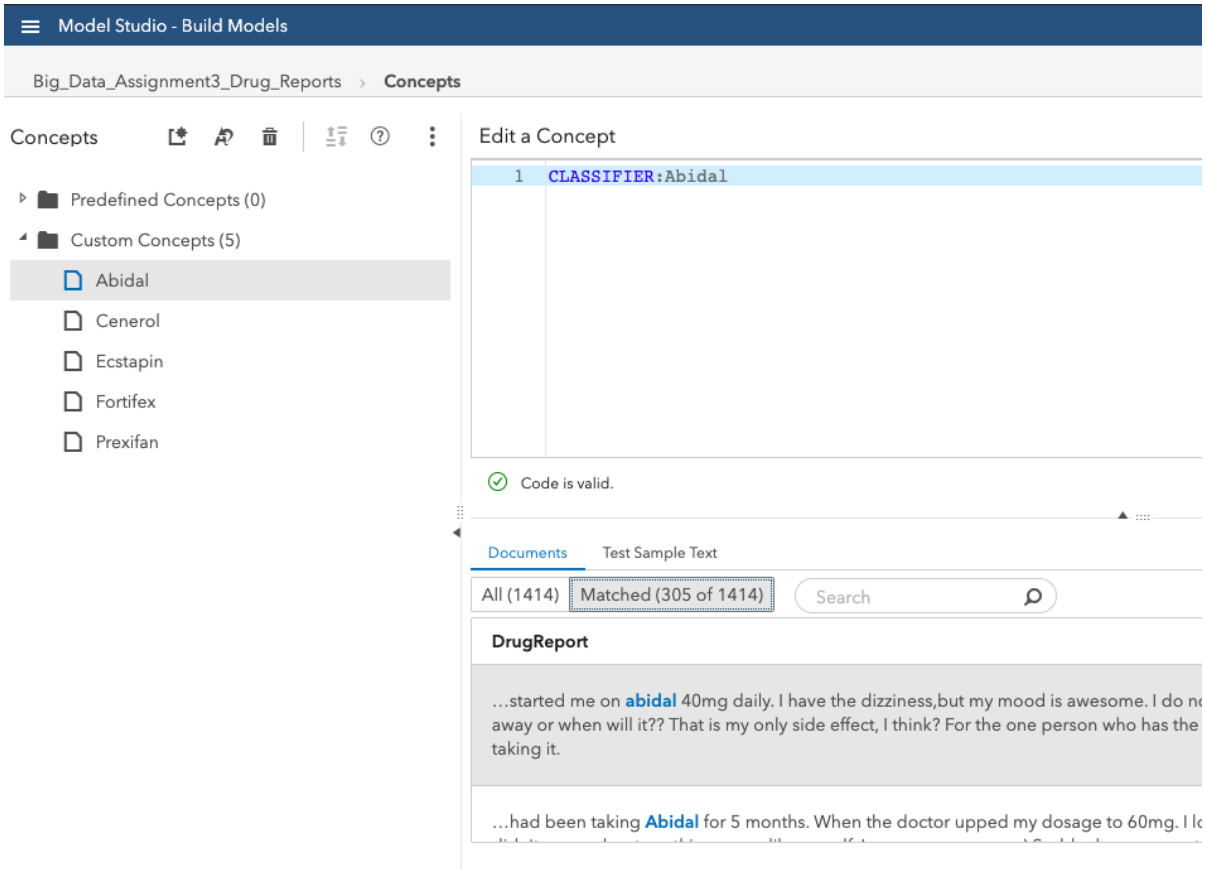
Task 2 - How many documents mention each medication

Display the counts of number of documents that mention each of the following medications:

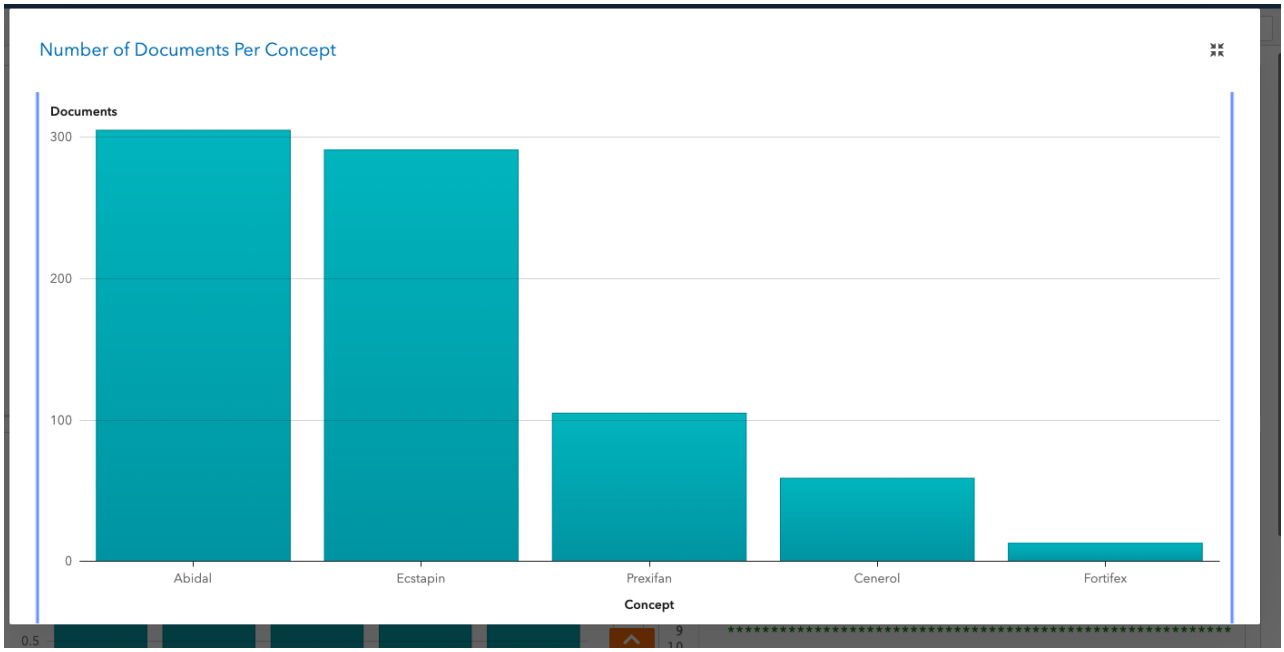
1. **Abidal**
2. **Cenerol**
3. **Ecstapin**
4. **Fortifex**
5. **Prexifan**

For Task 2, I used only SAS Viya for Learners.

So, I created 5 different concepts (Abidal, Cenerol, Ecstapin, Fortifex, & Prexifan) and pasted the required classifiers in all the 5 concepts from the given Medication text file. In addition, I validated the rules and ran node for every concept.

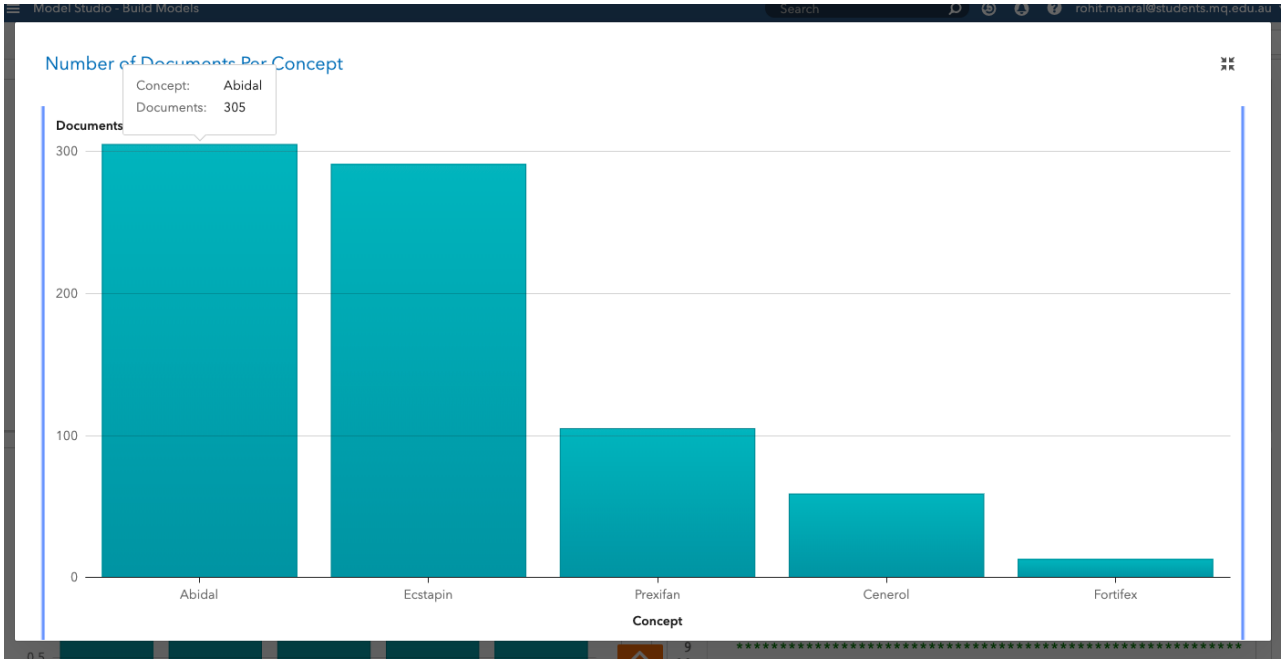


Then, we will go to *Pipeline* again and *run* Concepts. Then, select *Results* for Concepts.
Finally, we will get the bar graph representing the count for Abidal, Cenerol, Ecstapin, Fortifex, & Prexifan.

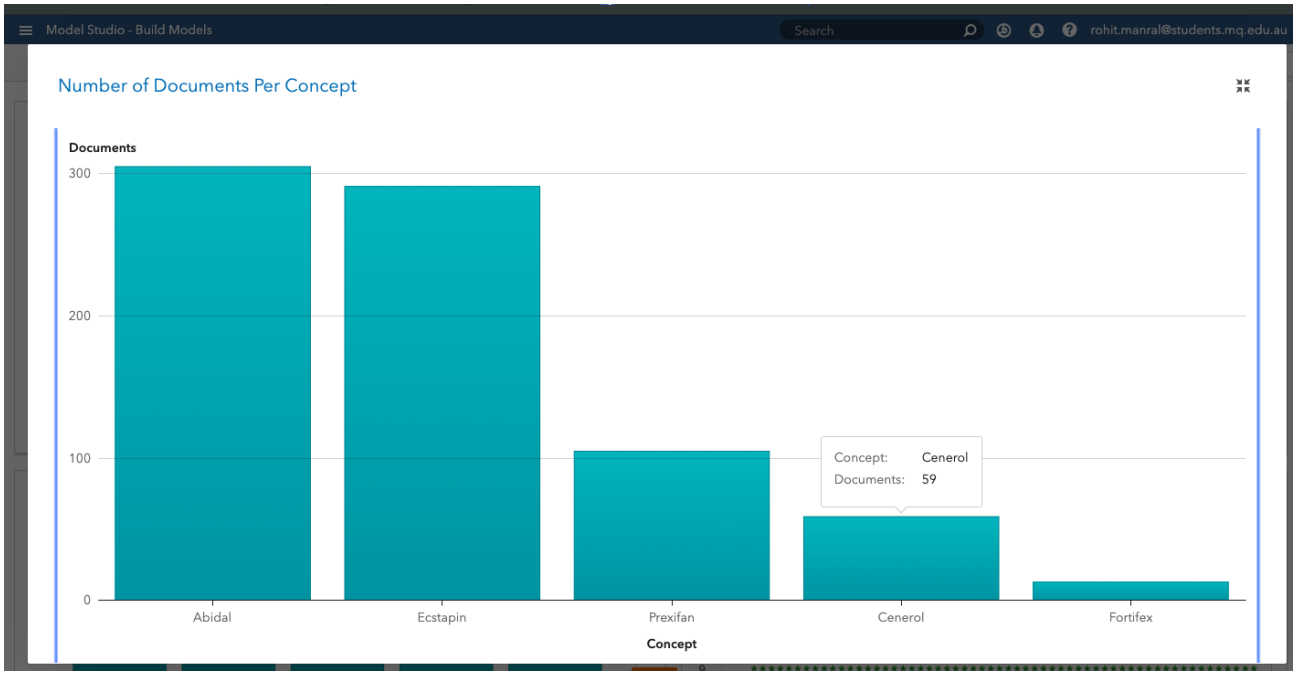


By scrolling or hovering over the different bars we can get the count of each element i.e. Abidal, Cenerol, Ecstapin, Fortifex, & Prexifan.

- **Scrolling or hovering over Abidal Bar**



- **Scrolling or hovering over Cenerol Bar**



- **Scrolling or hovering over Ecstapin Bar**

- **Scrolling or hovering over Fortifex Bar**

- **Scrolling or hovering over Prexifan Bar**

Therefore, ***305*** documents mention ***Abidal***, ***59*** documents mention ***Cenerol***, ***291*** documents mention ***Ecstapin***, ***13*** documents mention ***Fortifex*** & ***105*** documents mention ***Prexifan***.

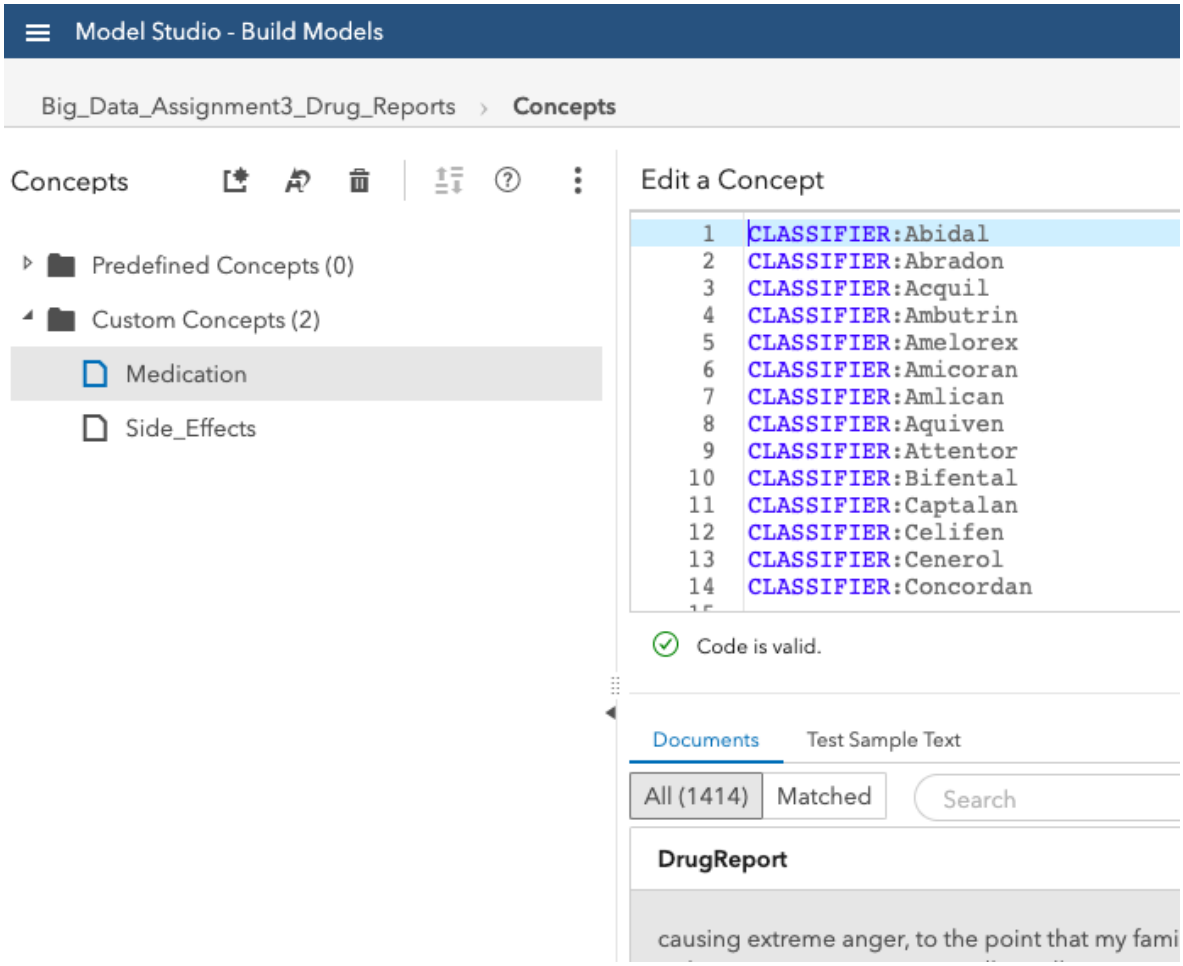
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Task 3 (5 marks) - Which medications are associated with sleep issues?

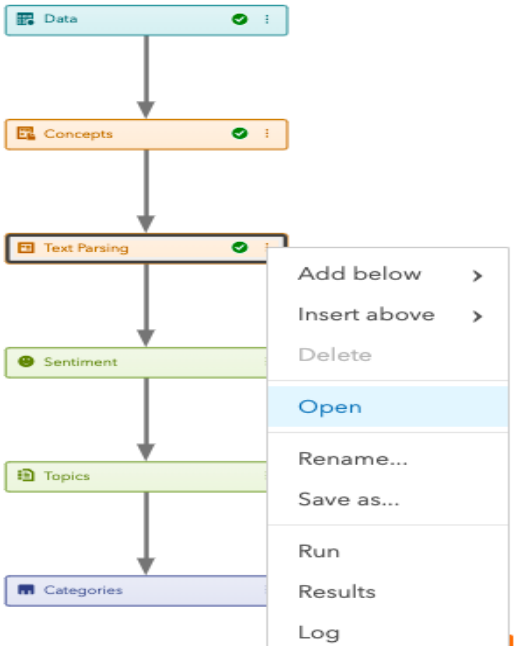
Identify the medications that are associated with sleep issues. Your chart or charts should clearly indicate which medications have stronger associations with sleep issues, and they should quantify the degree of association.

For Task 3, I used only SAS Viya for Learners.

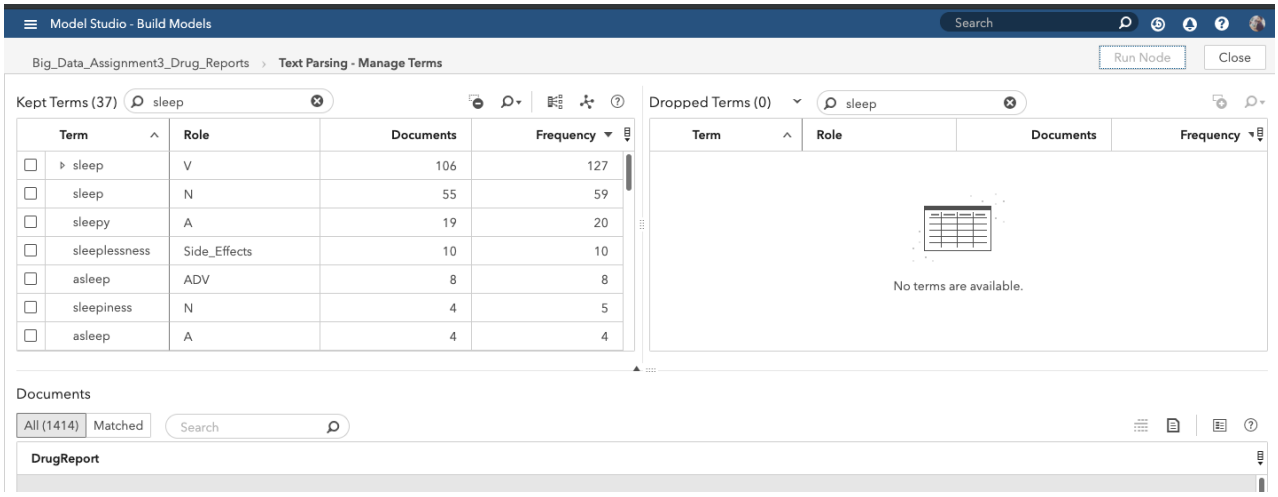
So, I made 2 Custom Concepts i.e. Medication & Side Effects.



Then, opened Text Parsing.

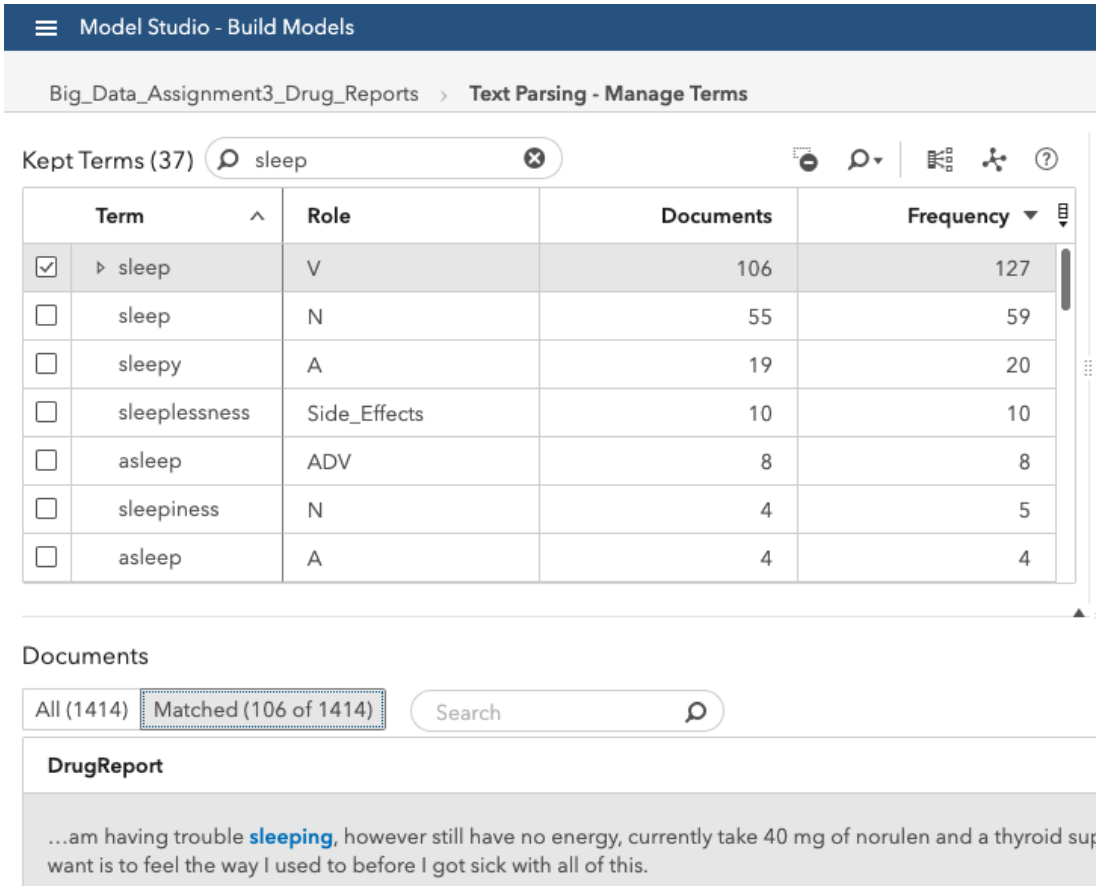


Then, searched for ***sleep*** in both the Kept terms and Dropped terms. In addition, added all the dropped terms in Kept terms.

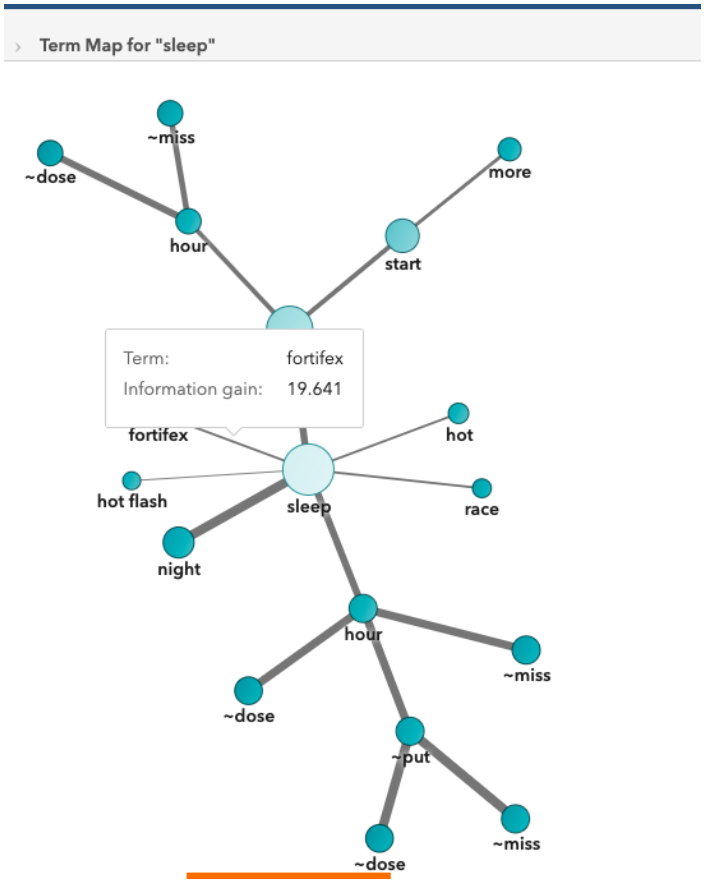


Finally, I started checking all the terms related to sleep one-by-one. Here, the word ‘sleep’ and its relevant words are present in many forms i.e. Verb, Noun, Adjective, Proper Noun, etc. So, I made a separate check for all of them using Term Map.

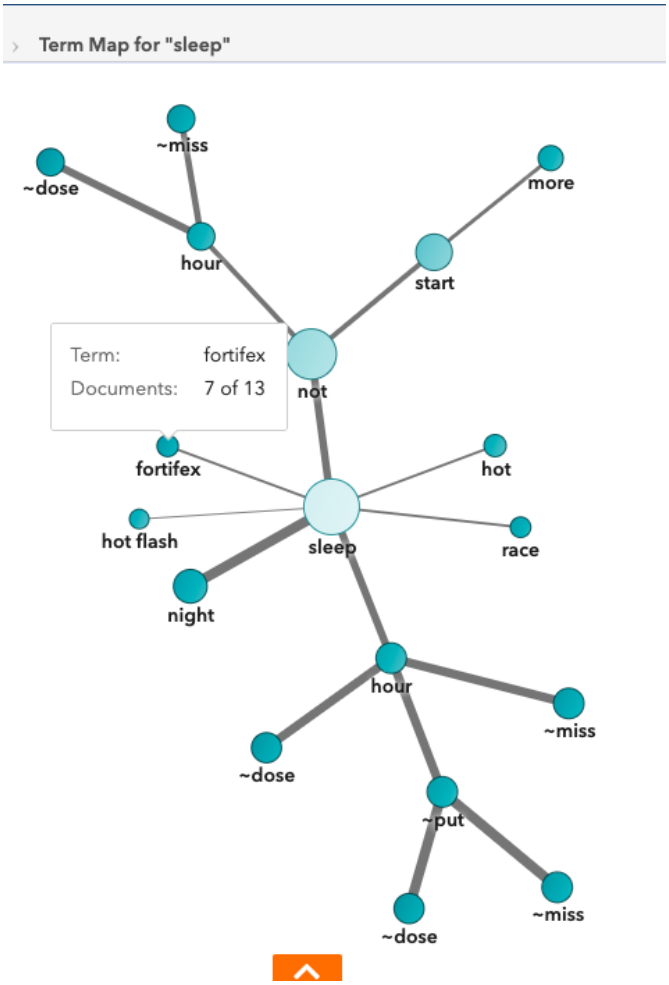
- ***sleep (Verb)***



There are 106 forms of ***sleep (Verb)*** present in 1414 documents.

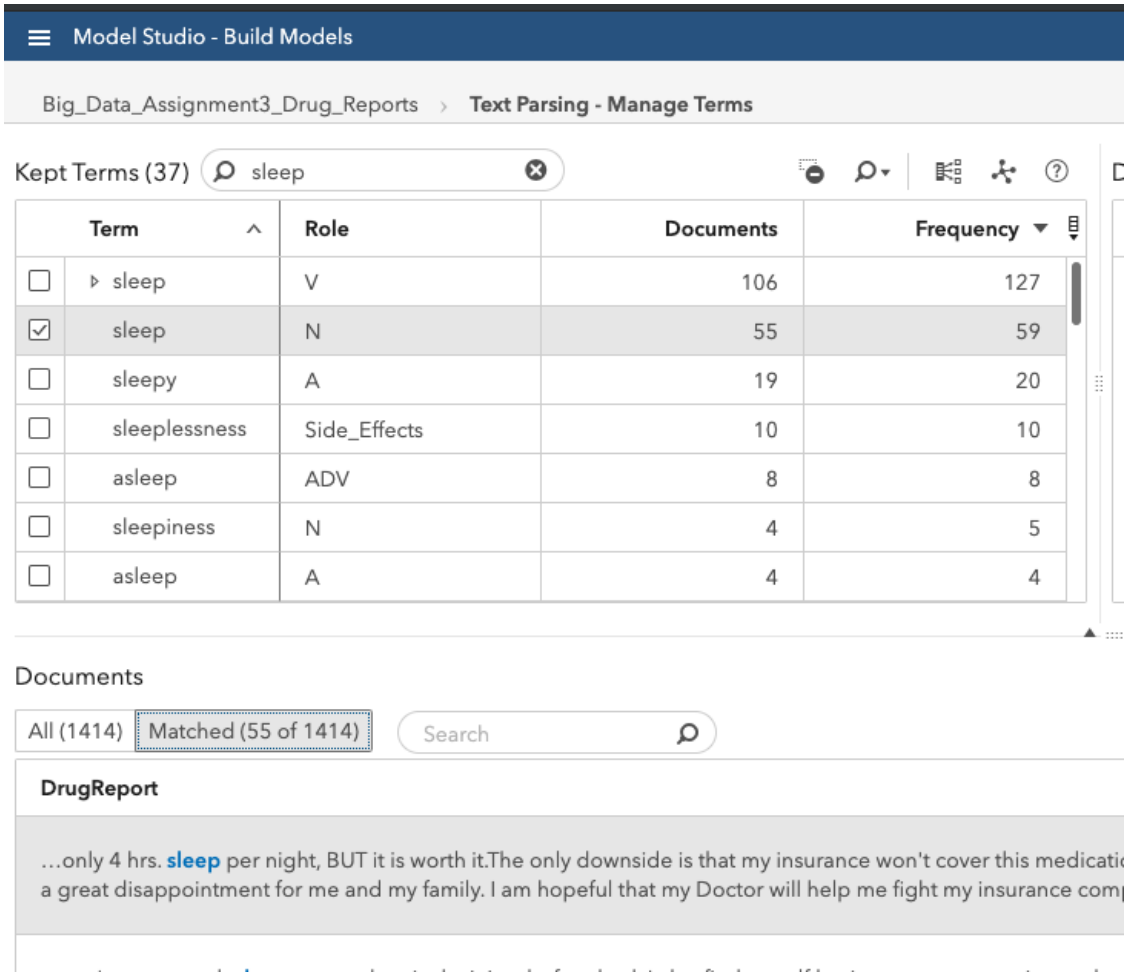


The Information gain between *sleep (Verb)* and fortifex is 19.641.

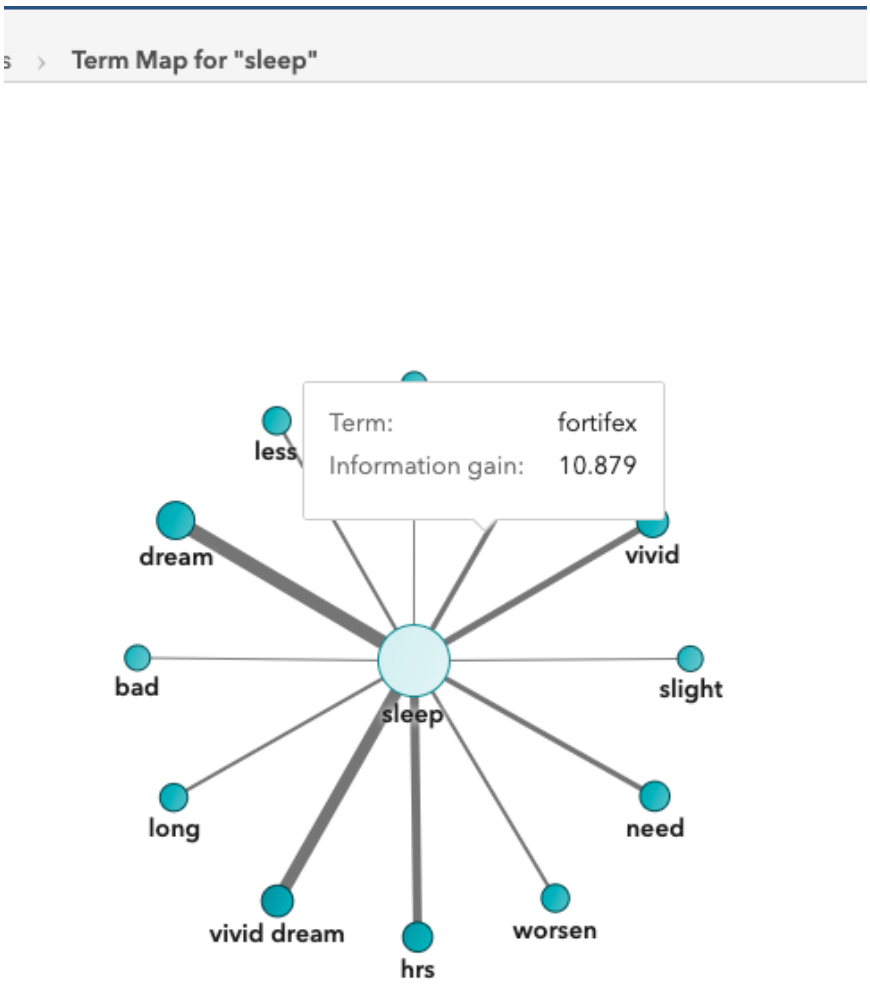


So, there are total 13 fortifex ***medications*** out of those 7 are ***associated with sleep (Verb)***.

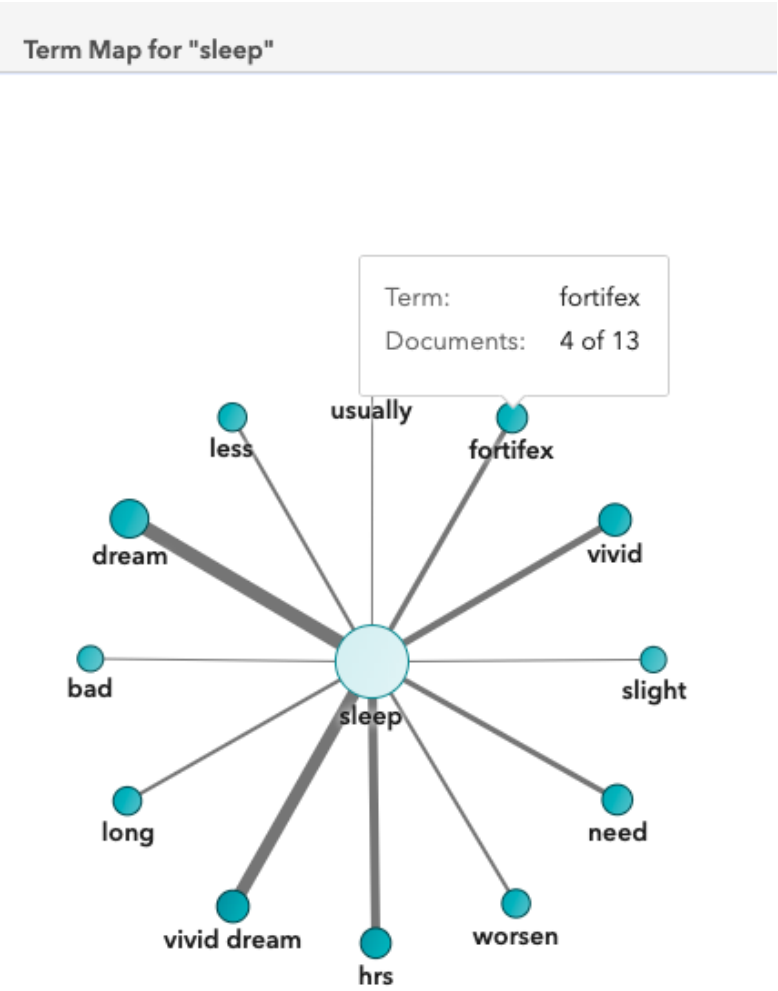
- ***sleep (Noun)***



There are 55 ‘sleep’ nouns in 1414 documents.

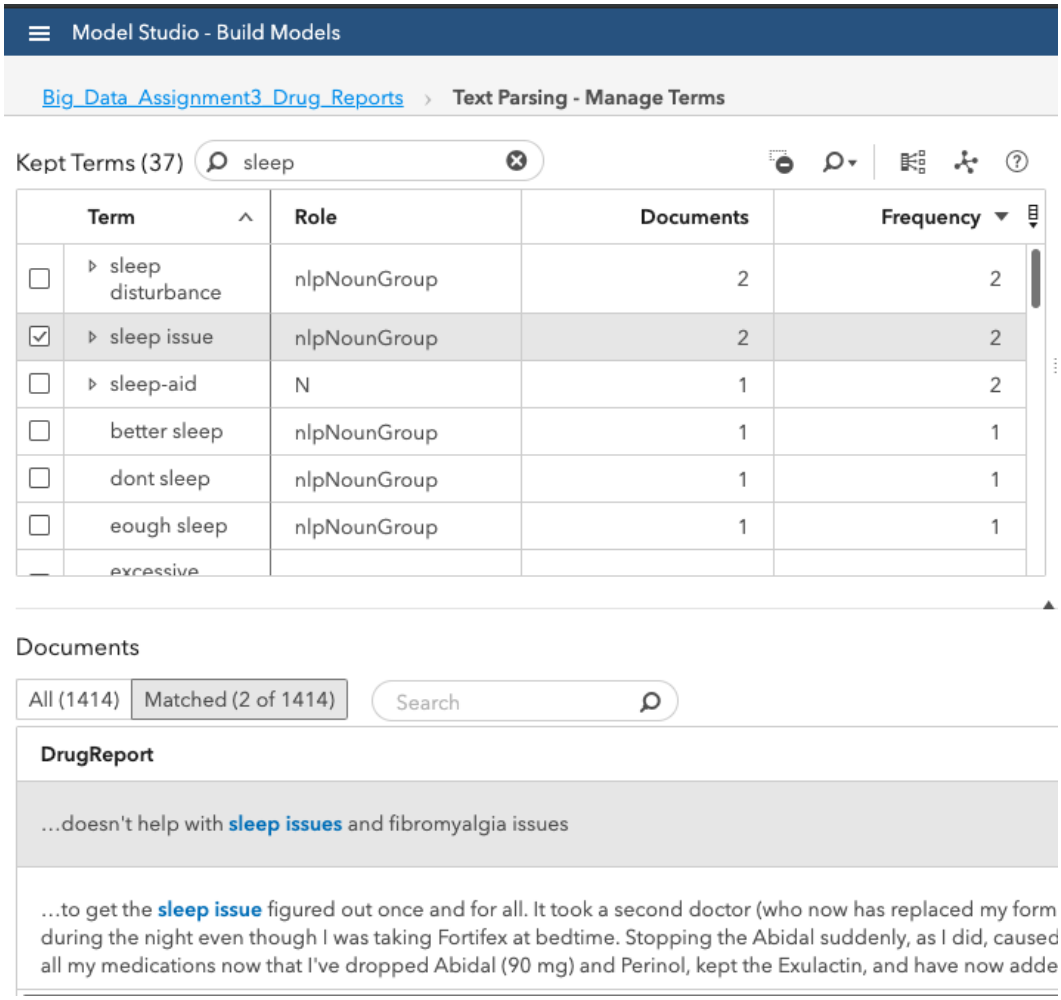


The Information gain between *sleep (Noun)* and fortifex is 10.879.

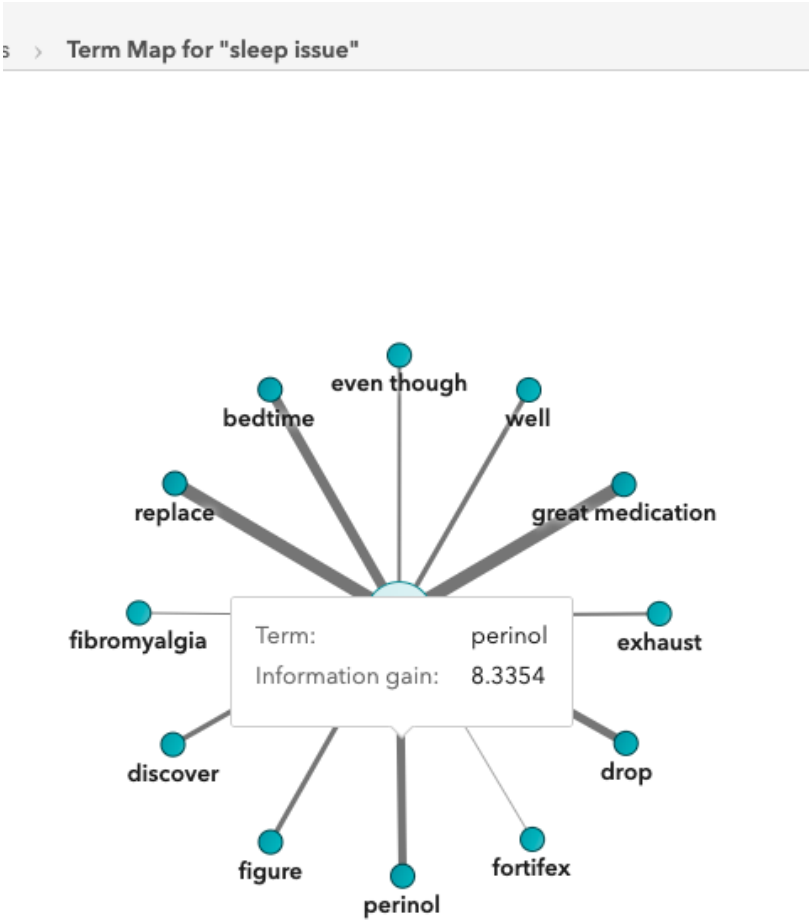


So, there are total 13 fortifex ***medications*** out of those 4 are ***associated with sleep (Noun)***.

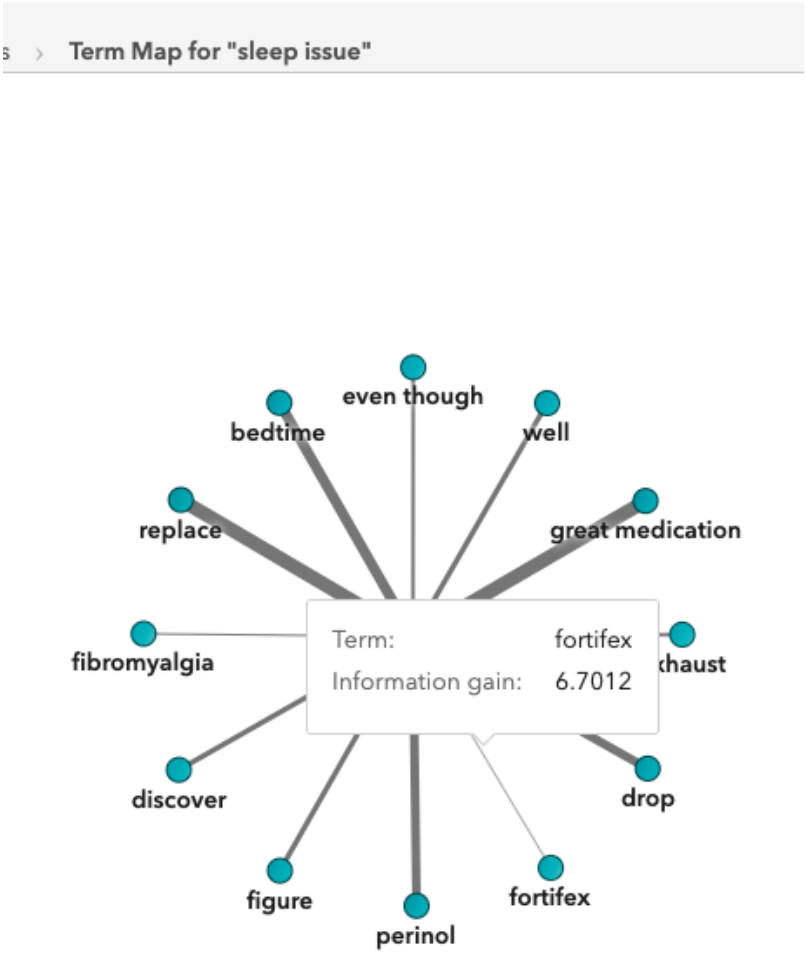
- ***sleep issue (nlpNounGroup)***



There are 2 ‘sleep issue’ nouns in 1414 documents.

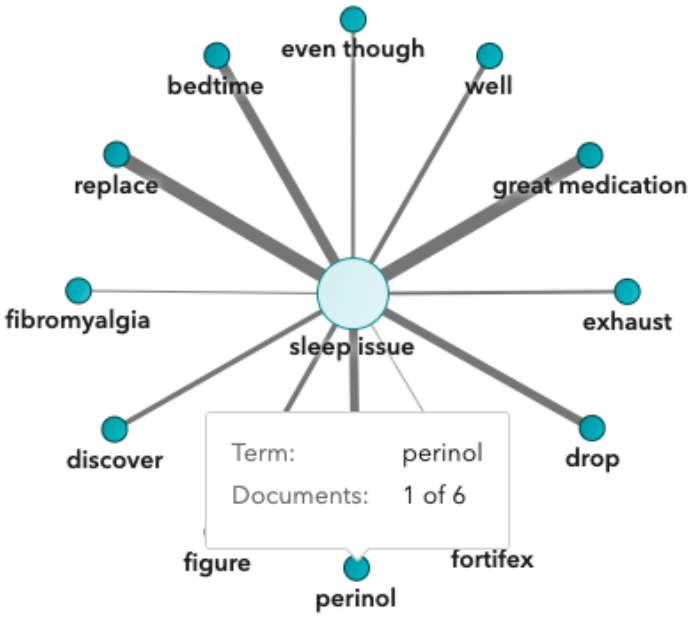


The Information gain between *sleep issue (Noun)* and perinol is 8.3354.

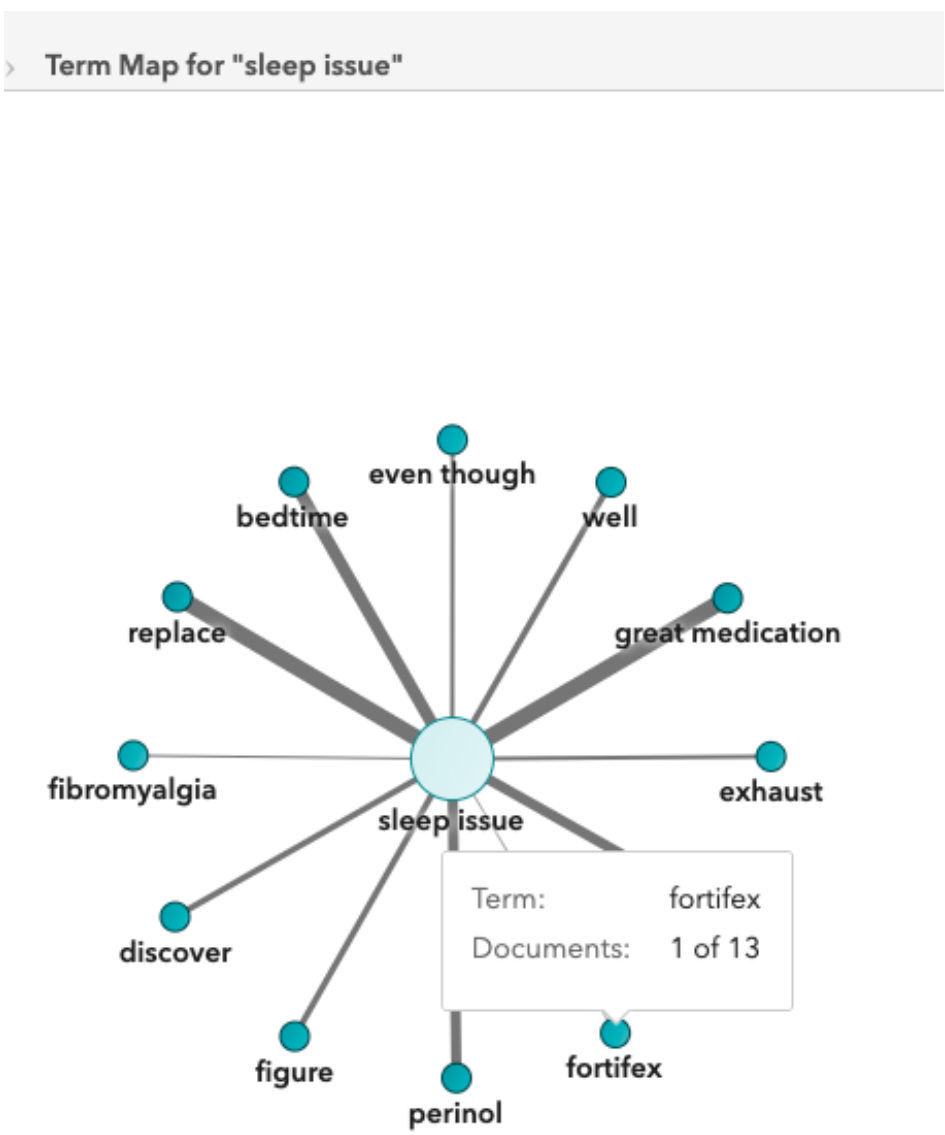


The Information gain between *sleep issue (Noun)* and fortifex is 6.7012.

Term Map for "sleep issue"

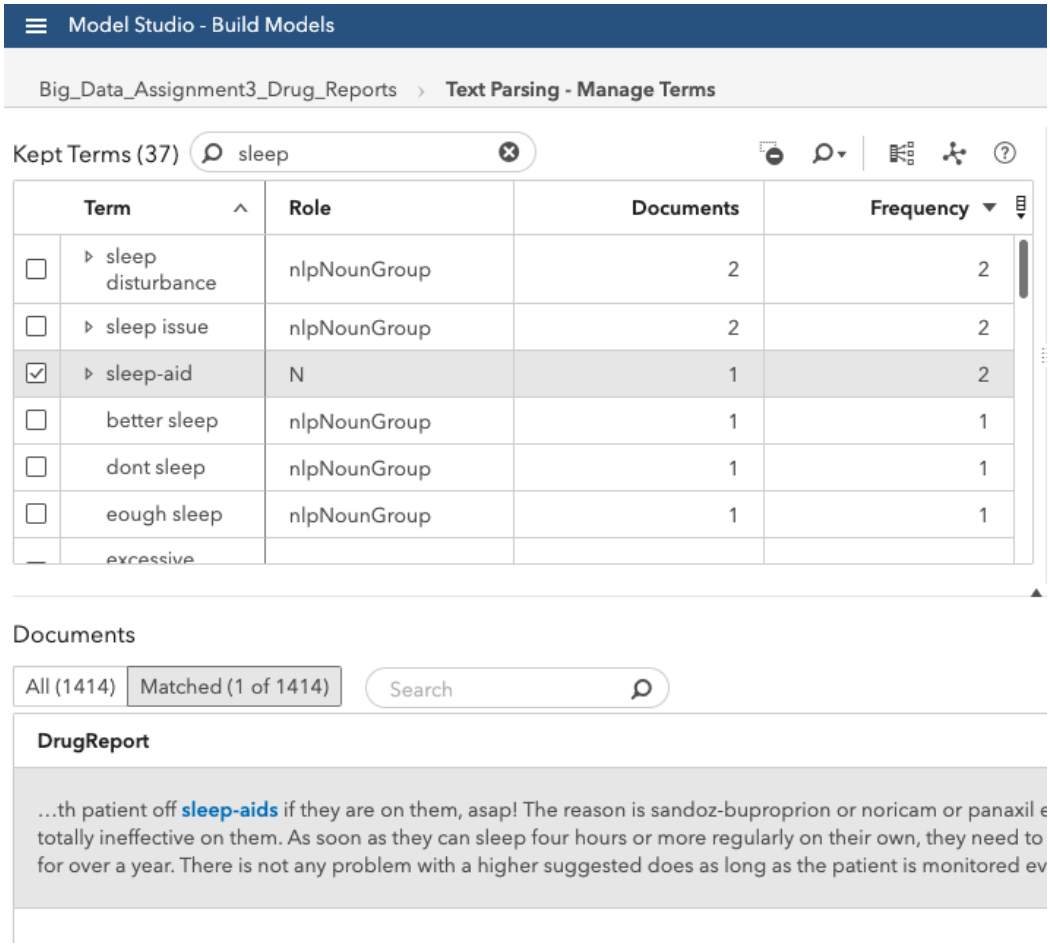


So, there are total 6 perinol ***medications*** out of those 1 is ***associated with sleep issue (Noun)***.

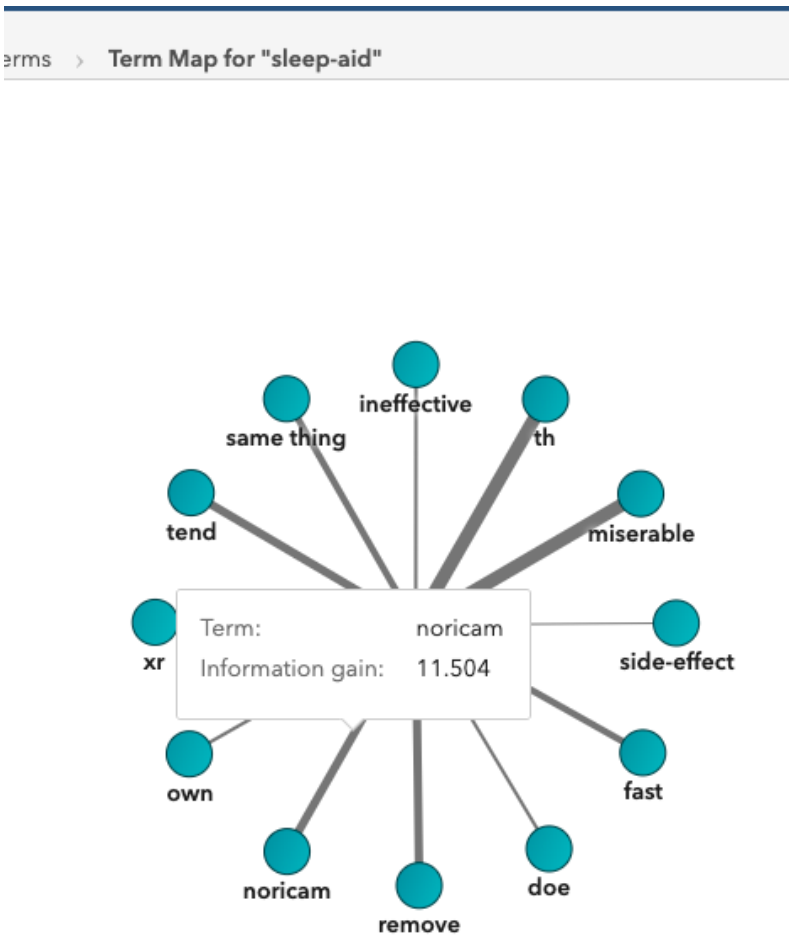


So, there are total 13 fortifex ***medications*** out of those 1 is ***associated with sleep issue (Noun)***.

- ***sleep-aid (Noun)***

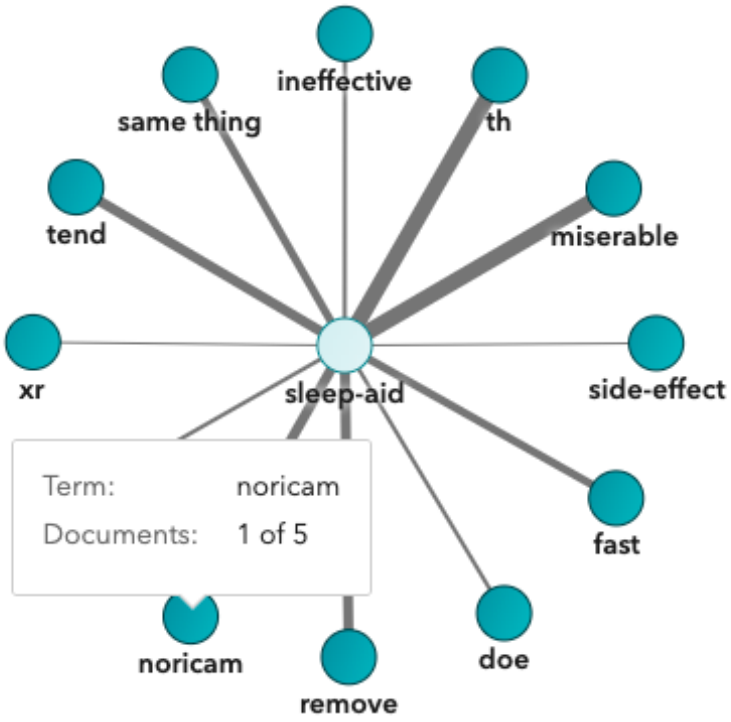


There is only one ‘sleep-aid’ noun in 1414 documents.



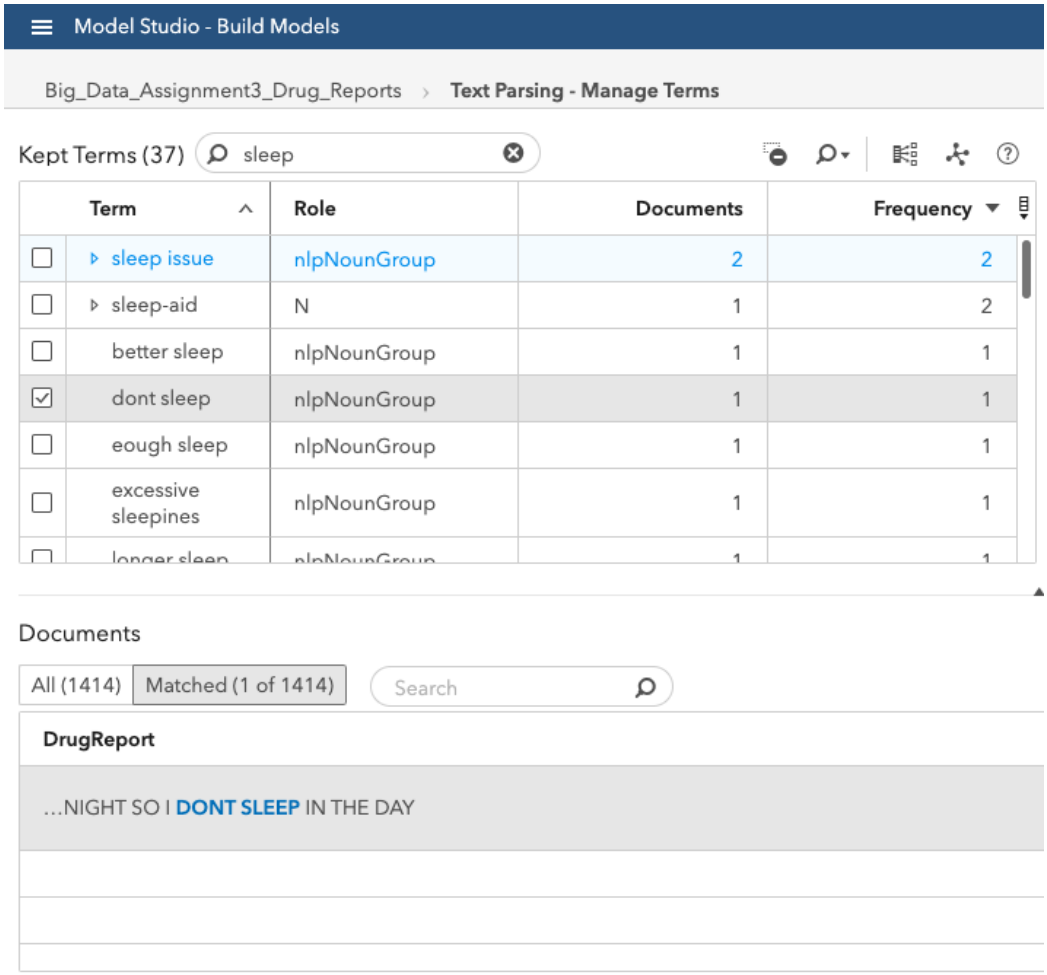
The Information gain between *sleep-aid (Noun)* and fortifex is 11.504.

Term Map for "sleep-aid"



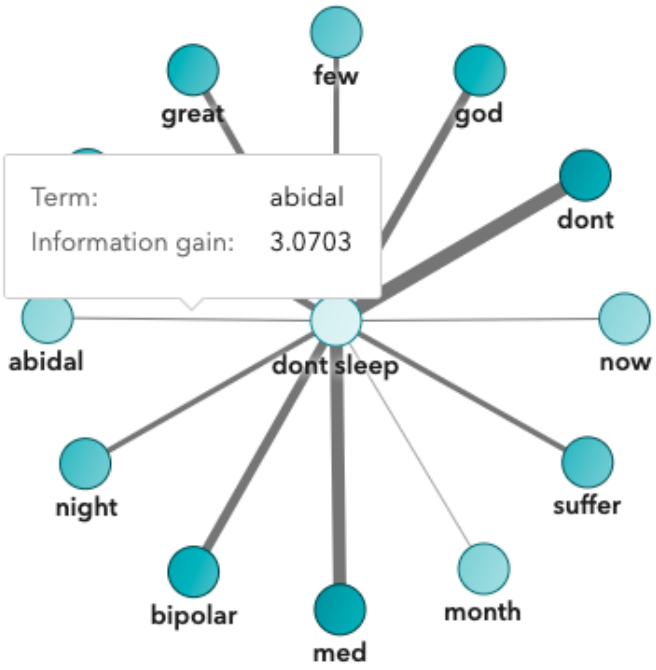
So, there are total noricam ***medications*** out of those 1 is ***associated with sleep-aid (Noun)***.

- ***dont sleep (nlpNounGroup)***

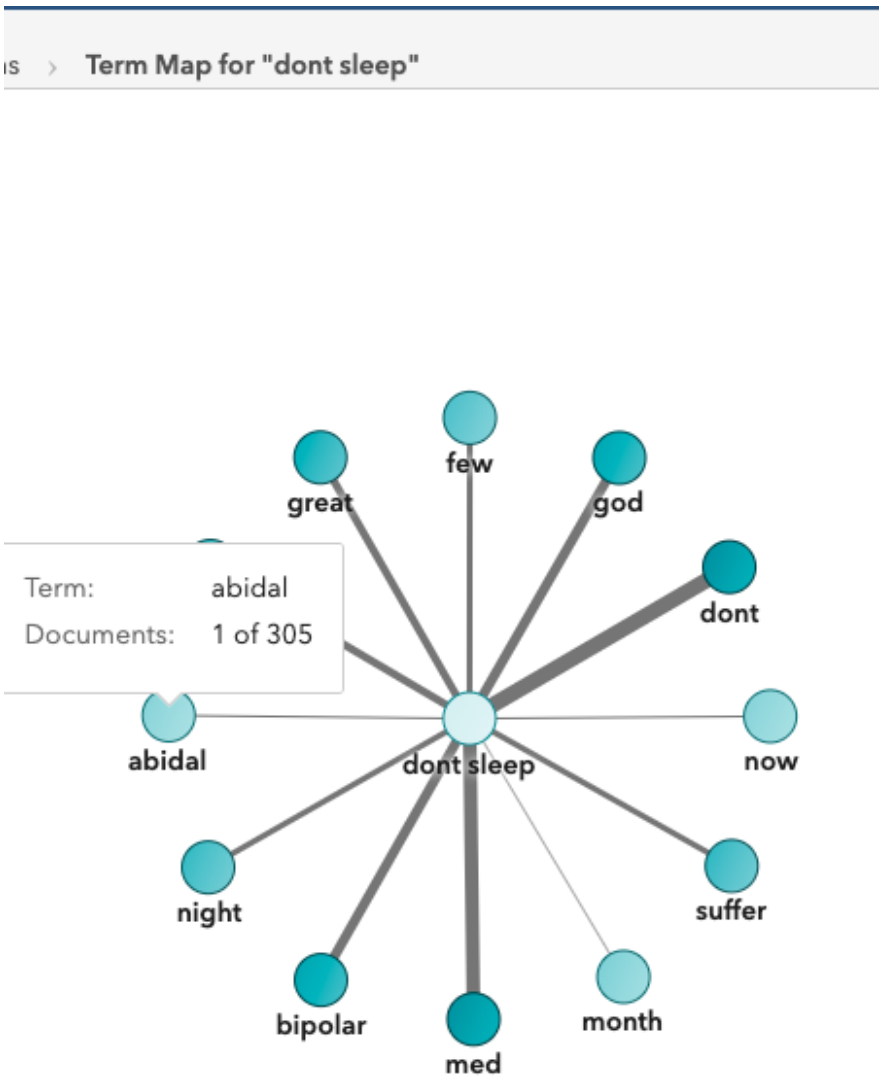


There is only one ‘dont sleep’ noun in 1414 documents.

Term Map for "dont sleep"

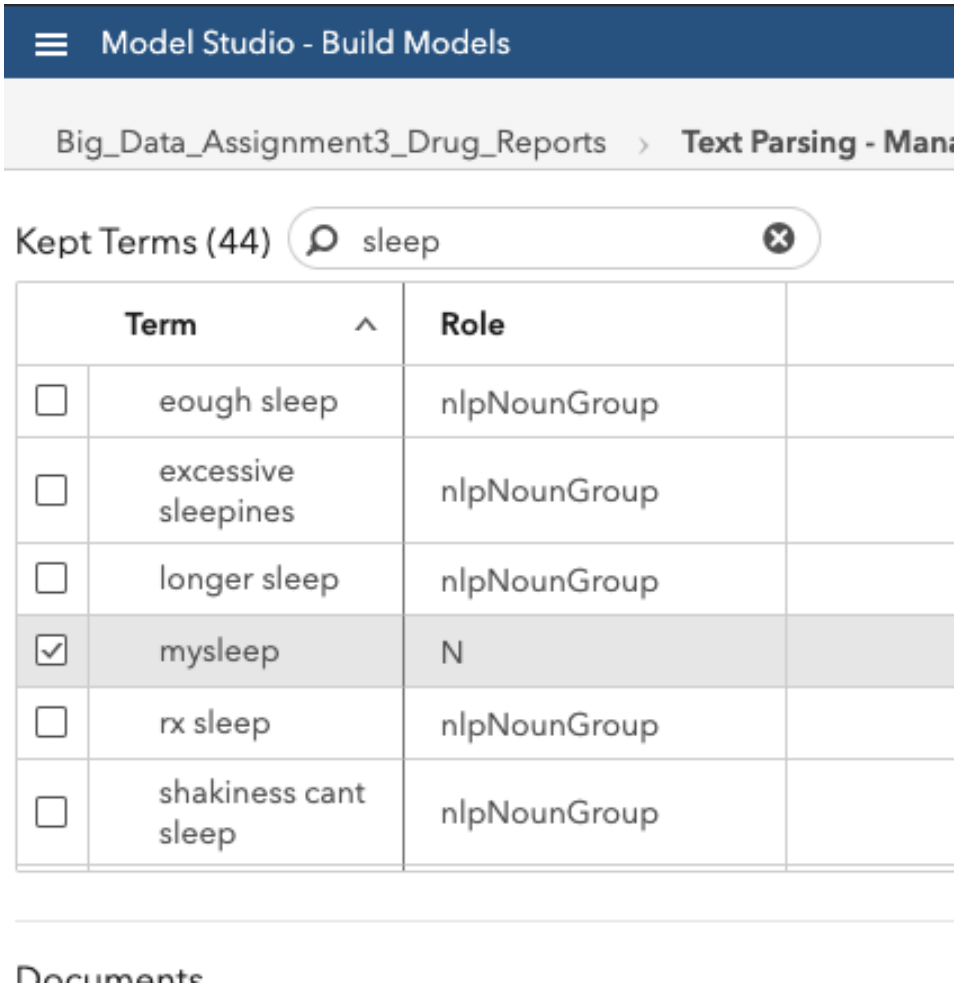


The Information gain between *dont sleep (Noun)* and abidal is 3.0703.



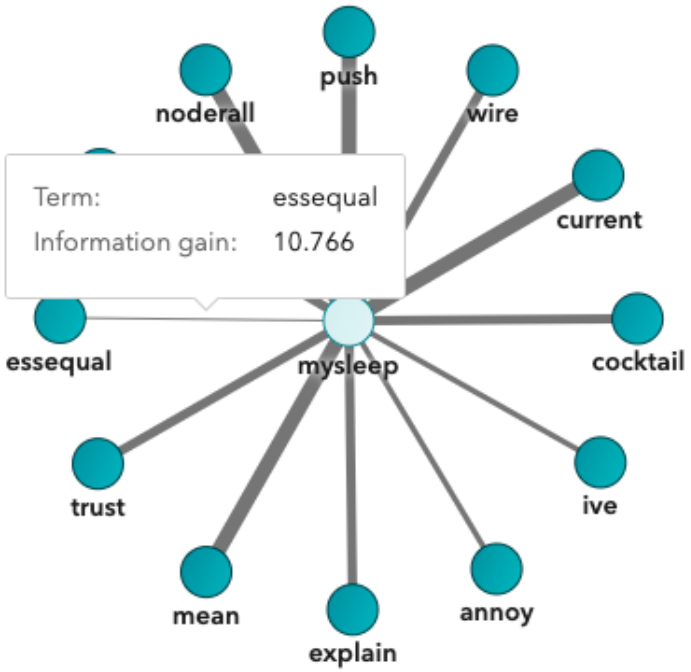
So, there are total 305 fortifex ***medications*** out of those 1 is ***associated with dont sleep (Noun)***.

- ***mysleep (Noun)***



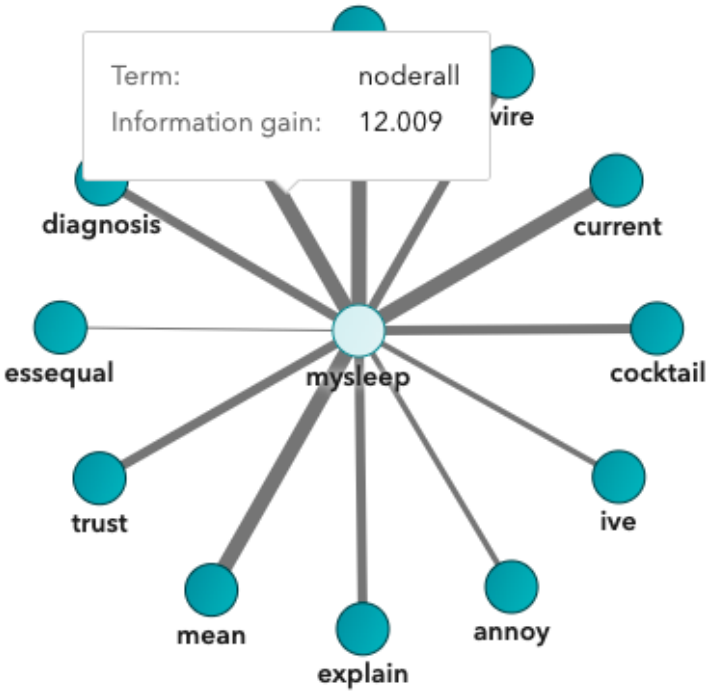
There is only one ‘mysleep’ noun in 1414 documents.

Term Map for "mysleep"

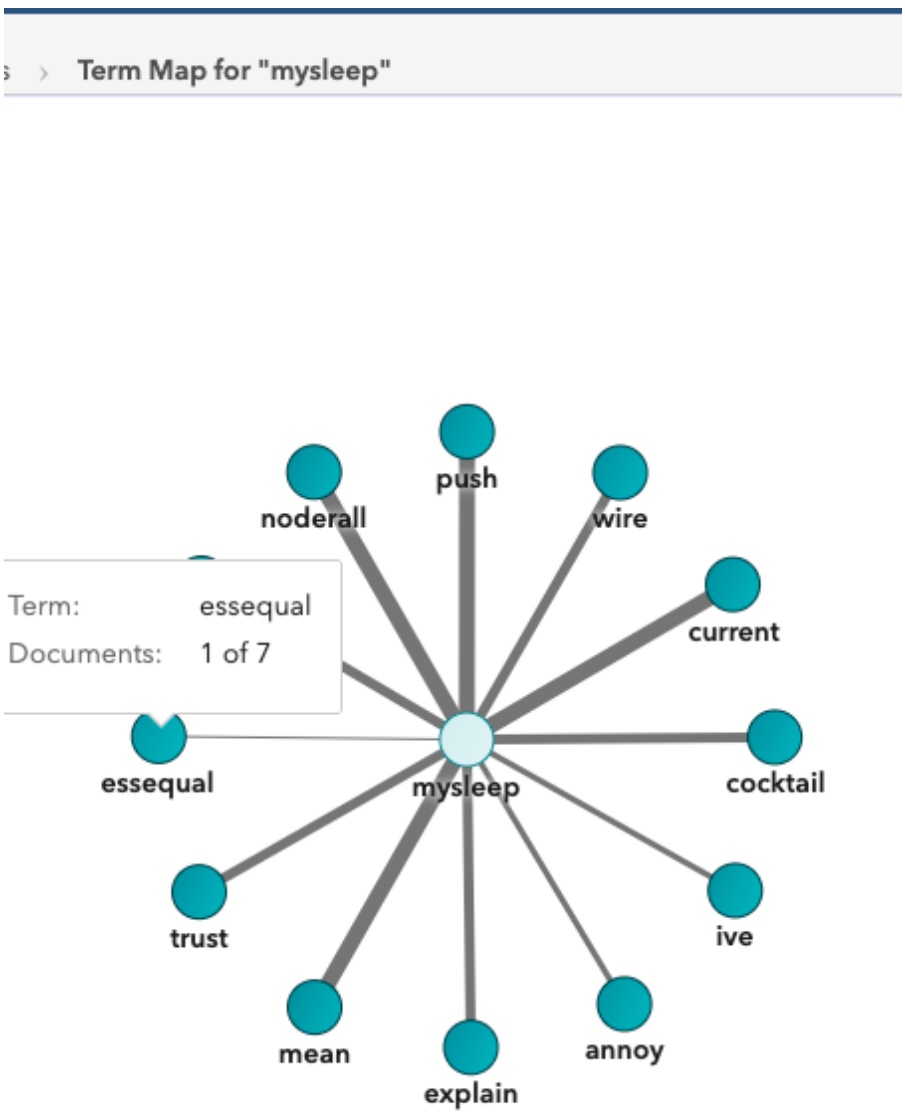


The Information gain between my*sleep (Noun)* and essequal is 10.766.

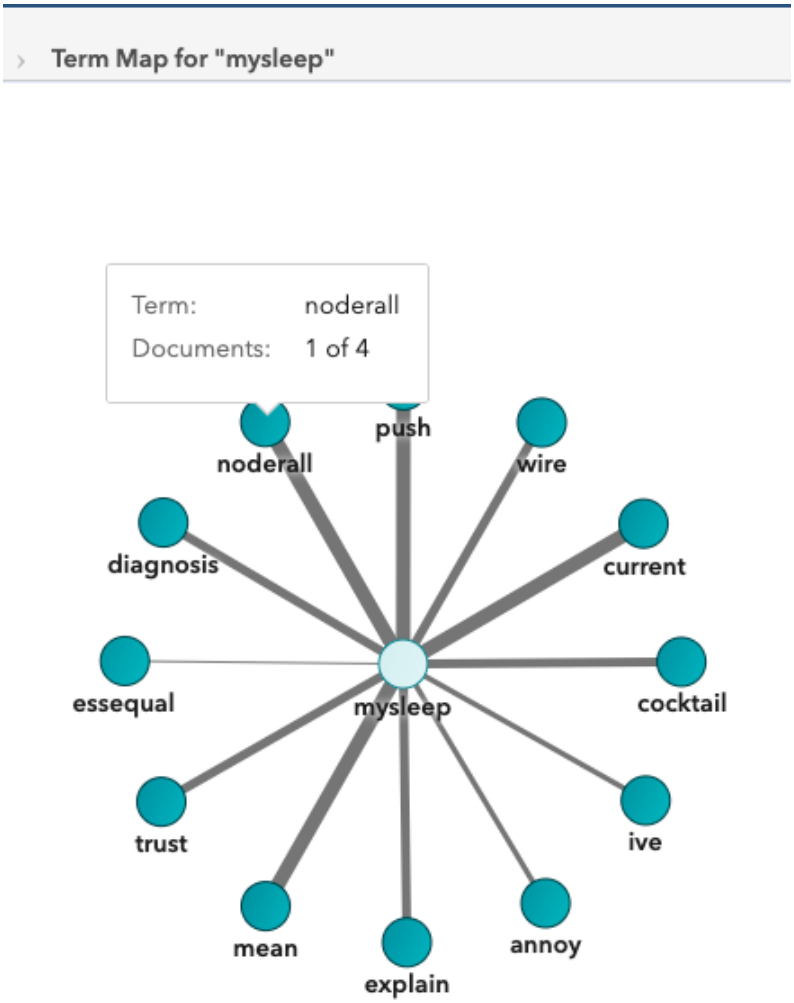
Term Map for "mysleep"



The Information gain between my*sleep (Noun)* and noderall is 12.009.

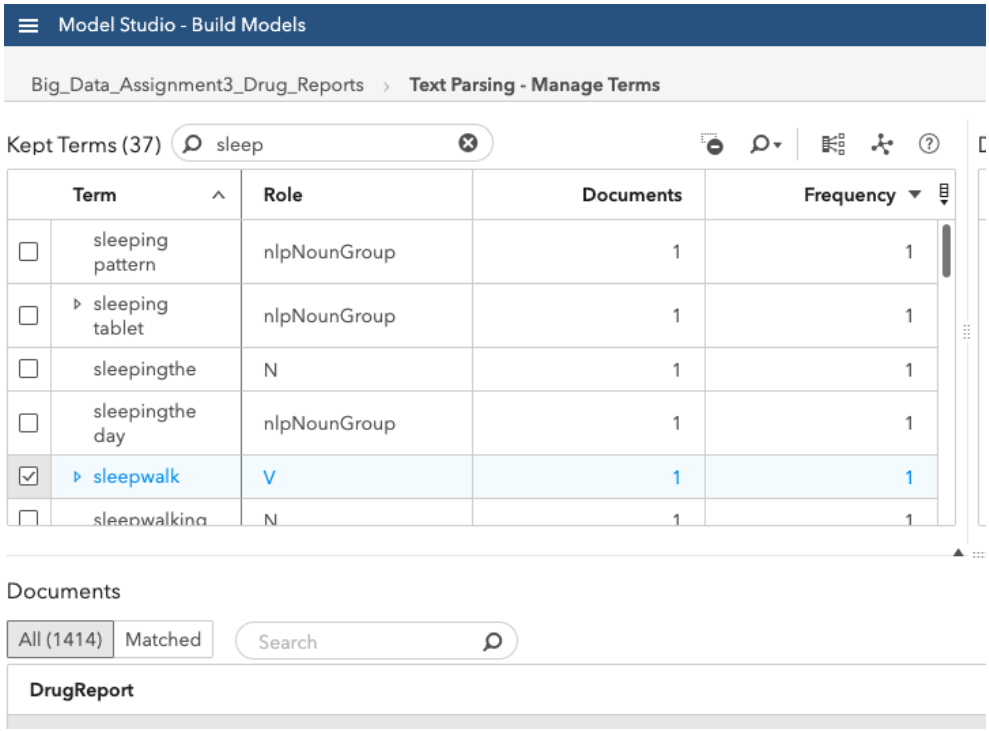


So, there are total 7 essequal ***medications*** out of those 1 is ***associated with mysleep (Noun)***.

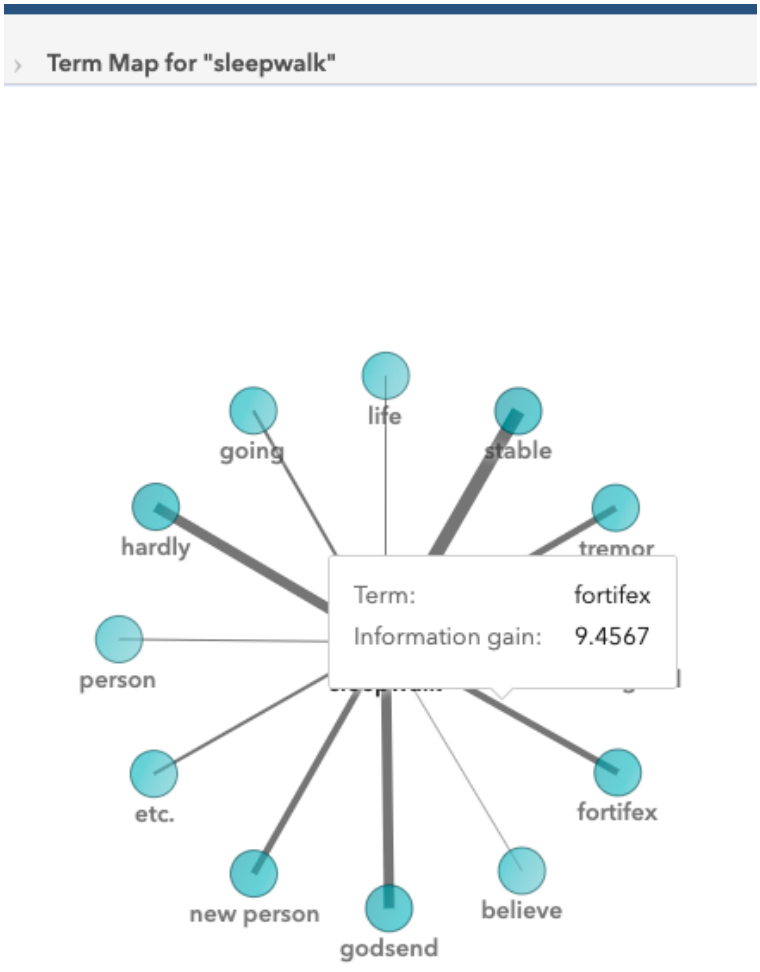


So, there are total 4 noderall ***medications*** out of those 1 is ***associated with mysleep (Noun)***.

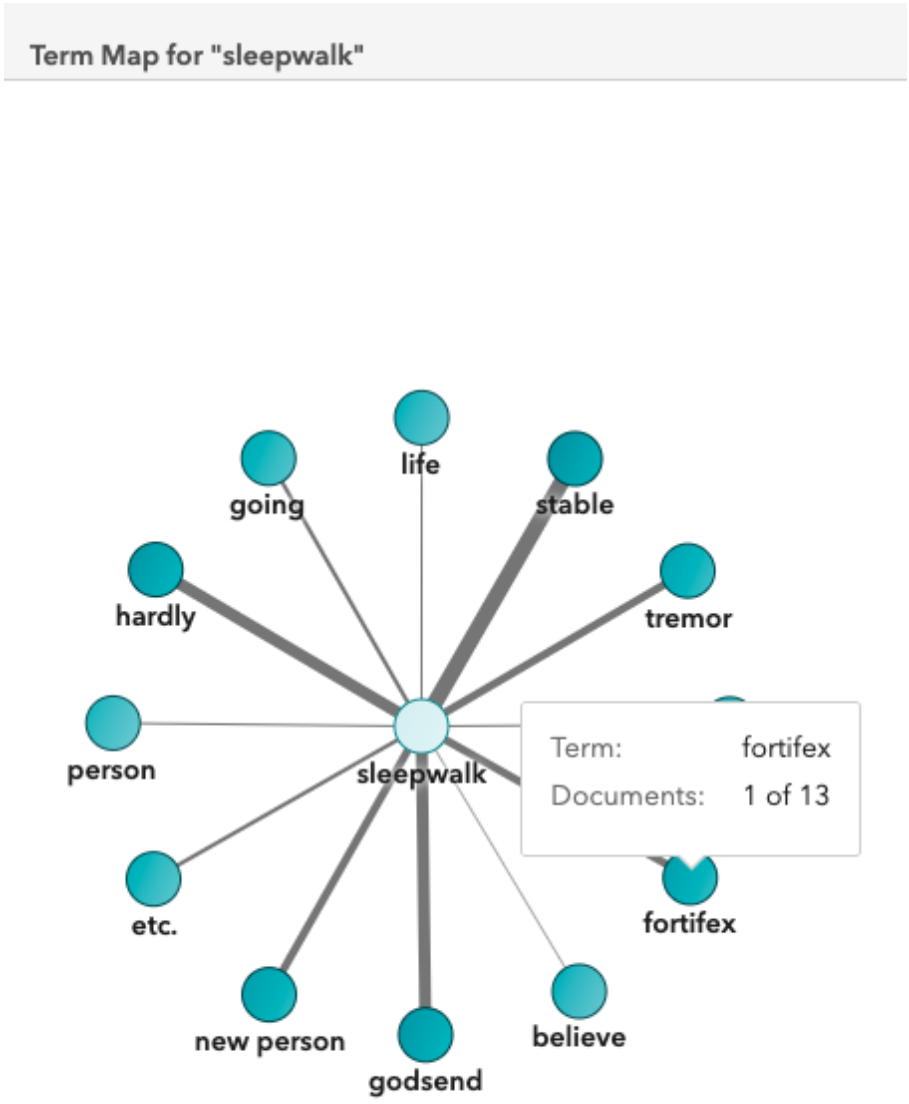
- ***sleepwalk (Verb)***



There is only one ‘sleepwalk’ verb in 1414 documents.

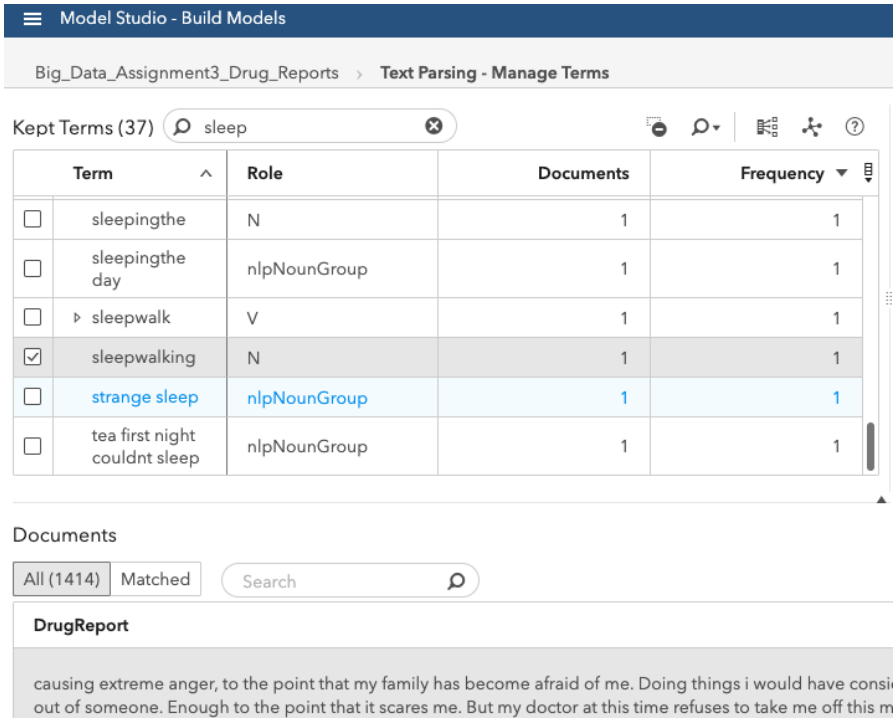


The Information gain between *sleepwalk (Verb)* and fortifex is 9.4567.



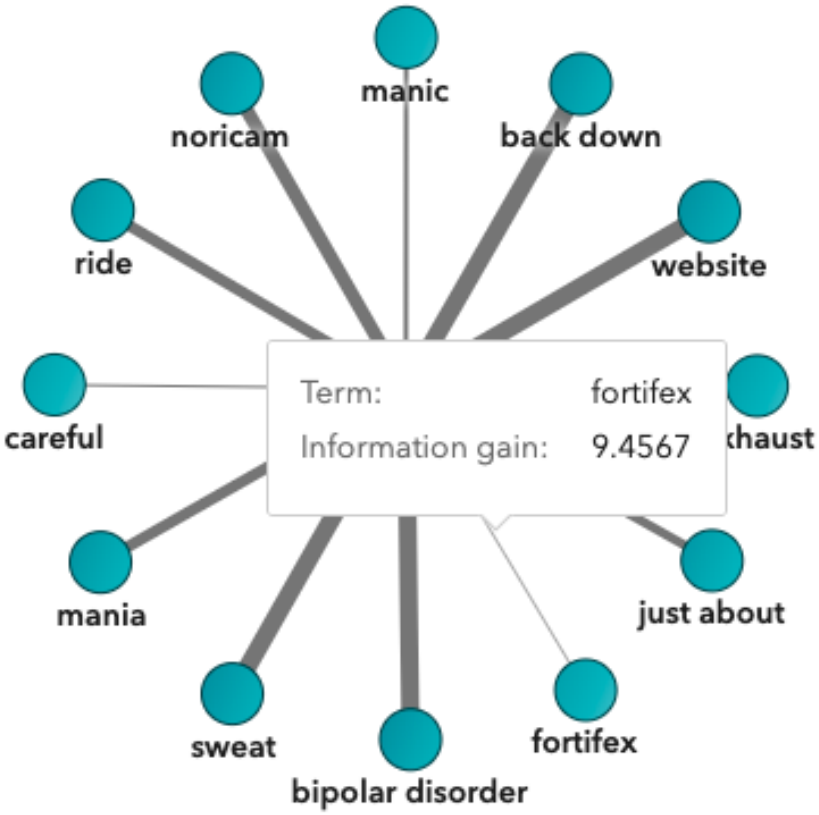
So, there are total 13 fortifex ***medications*** out of those 1 is ***associated with sleepwalk (Verb)***.

- ***sleepwalking (Noun)***



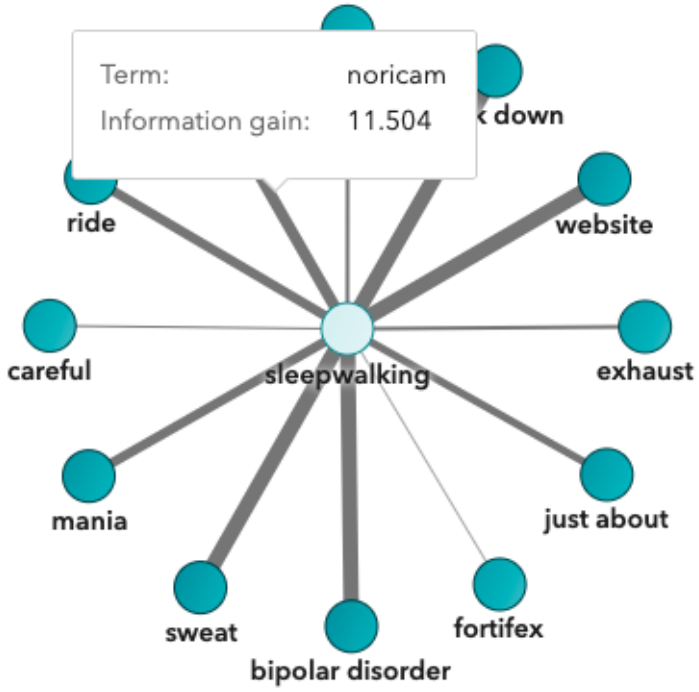
There is only one ‘sleepwalking’ noun in 1414 documents.

Term Map for "sleepwalking"

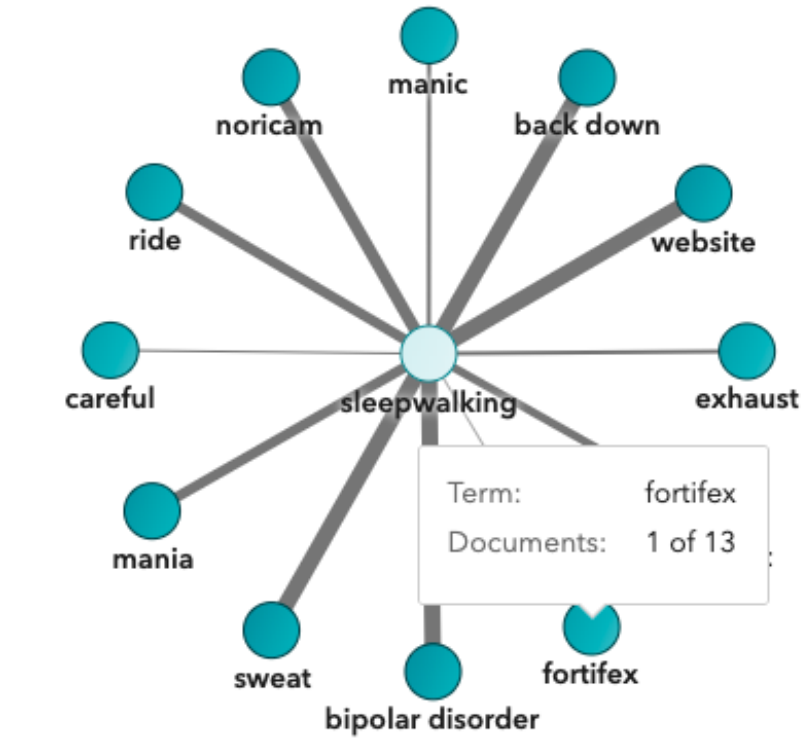


The Information gain between *sleepwalking (Noun)* and fortifex is 9.4567.

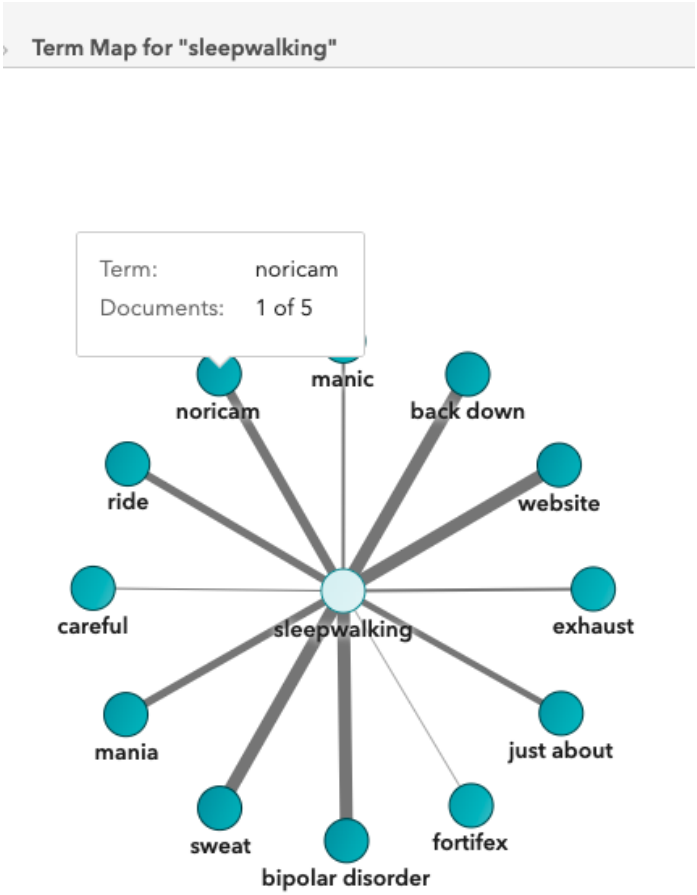
Term Map for "sleepwalking"



The Information gain between *sleepwalking (Noun)* and noricam is 11.504.



So, there are total 13 fortifex ***medications*** out of those 1 is ***associated with sleepwalking (Noun)***.



So, there are total 5 noricam ***medications*** out of those 1 is ***associated with sleepwalking (Noun)***

Therefore, the ***medications*** that are ***associated with sleep issues*** are ***Fortifex***, ***Noricam, Noderall, Essequal, Abidal, and Perinol***.

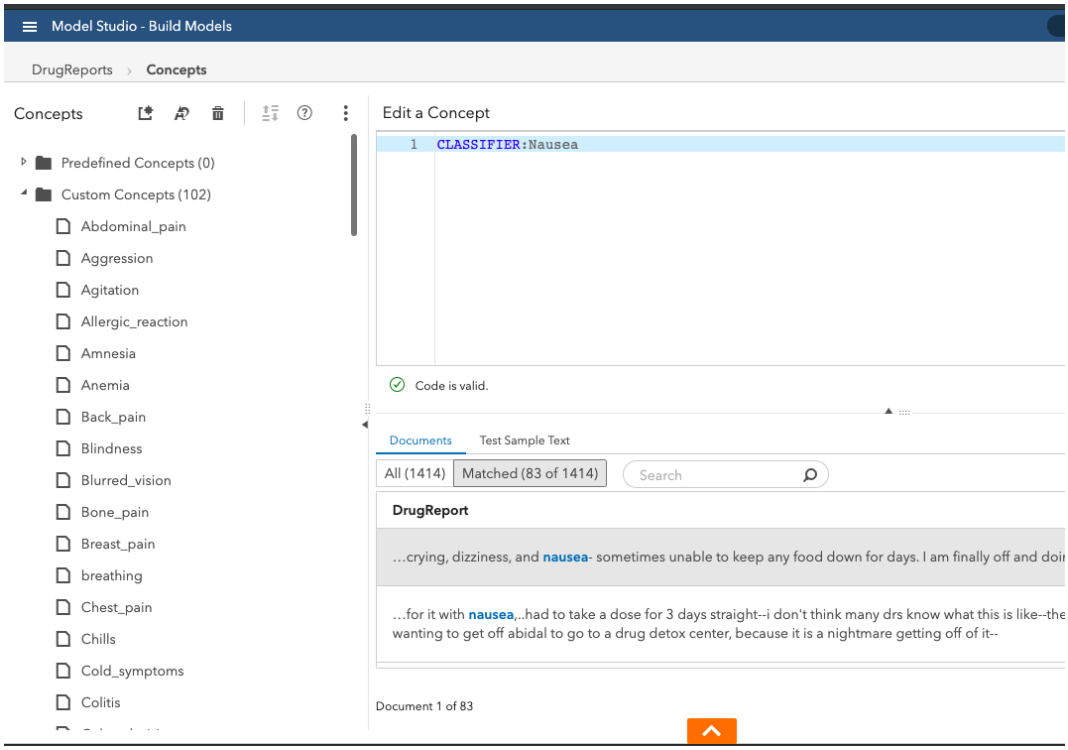
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Task 4 - What are the most common concerns expressed?

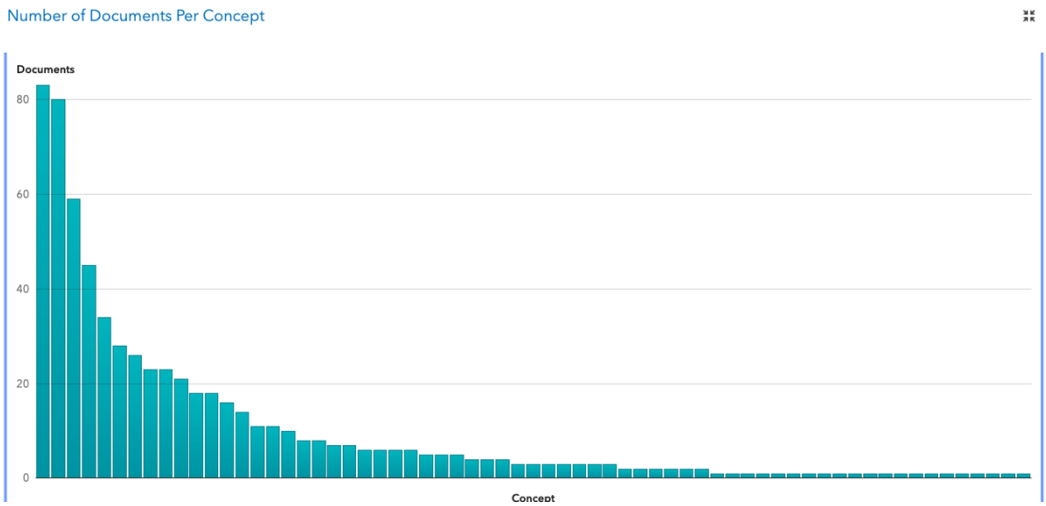
To complete this task, you need to determine the main concerns expressed in the posts. For this task you are free to use any tools or combination of tools. Make sure that you explain what tools you are using and how, and how to read the charts that you produce.

For Task 4, I used only SAS Viya for Learners.

So, I created 102 different concepts (Abdominal pain, Aggression, Agitation, Allergic reaction, Amnesia, and so on) as mentioned in SideEffects text file. Then, pasted the required classifiers in all the 102 concepts from the given SideEffects text file. In addition, I validated the rules and ran node for every concept.

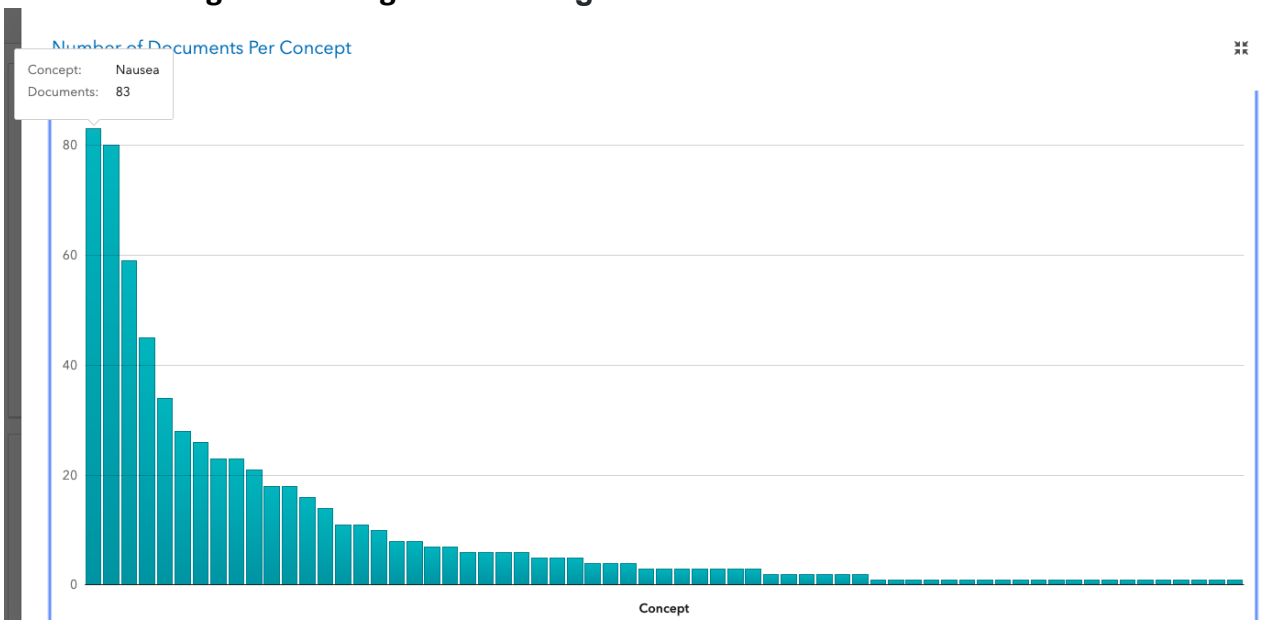


Then, we will go to *Pipeline* again and *run* Concepts. Then, select *Results* for Concepts.
Finally, we will get the bar graph representing the count for Abdominal pain, Aggression, Agitation, Allergic reaction, Amnesia, and so on.

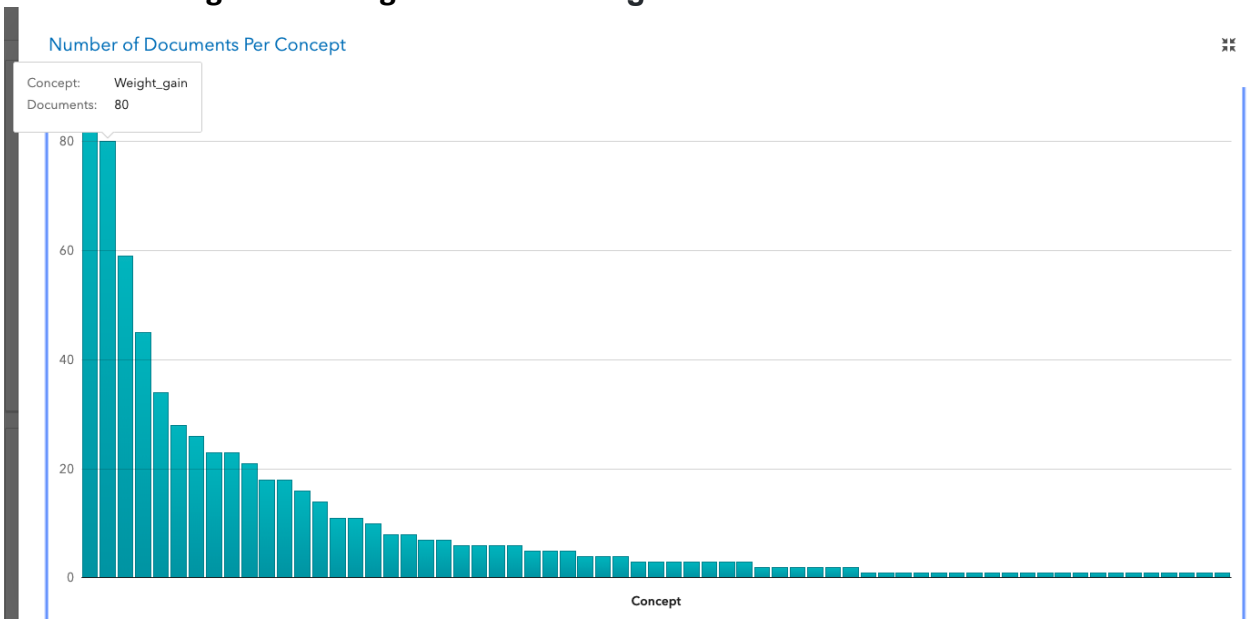


By scrolling or hovering over the different bars we can get the count of each element i.e. Abdominal pain, Aggression, Agitation, Allergic reaction, Amnesia, and so on.

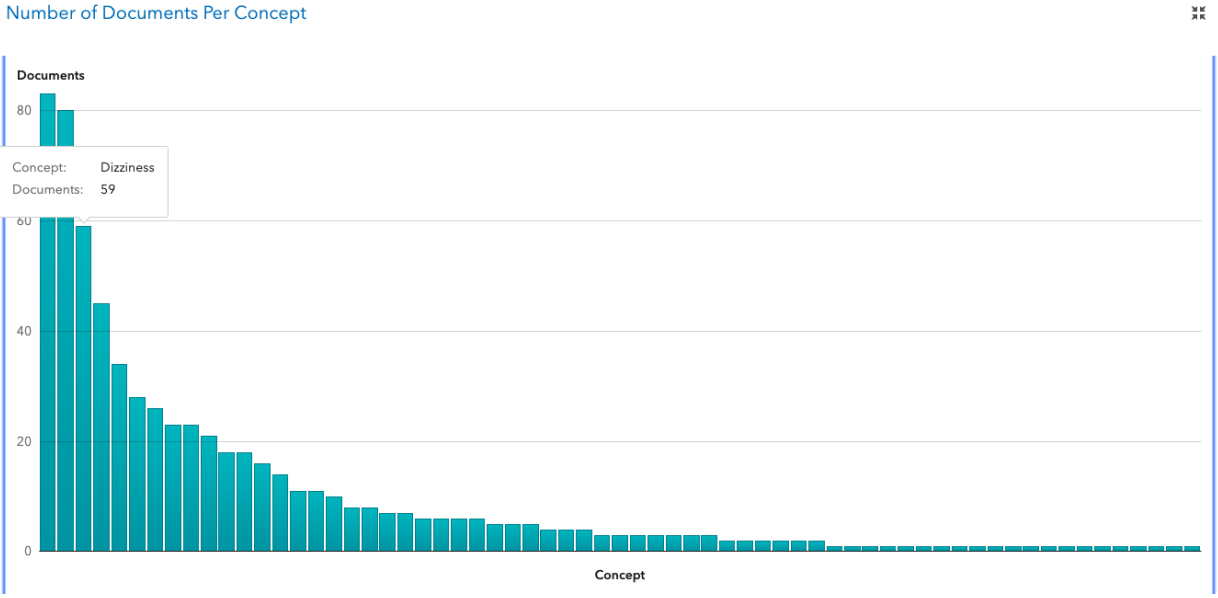
- **Scrolling or hovering over first highest Bar**



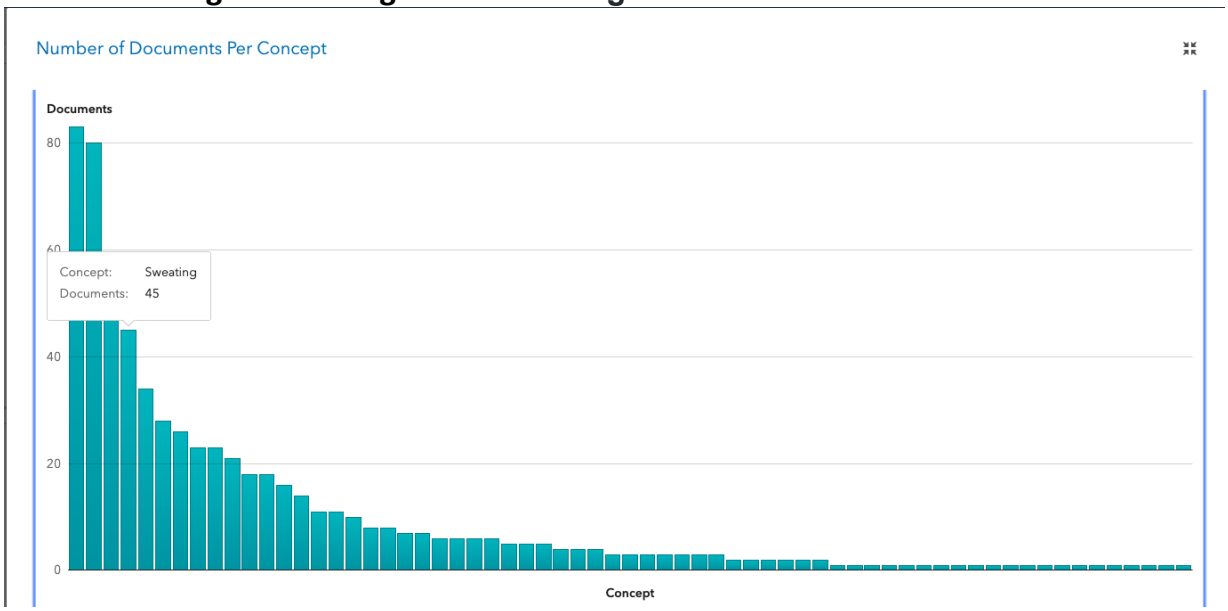
- **Scrolling or hovering over second highest Bar**



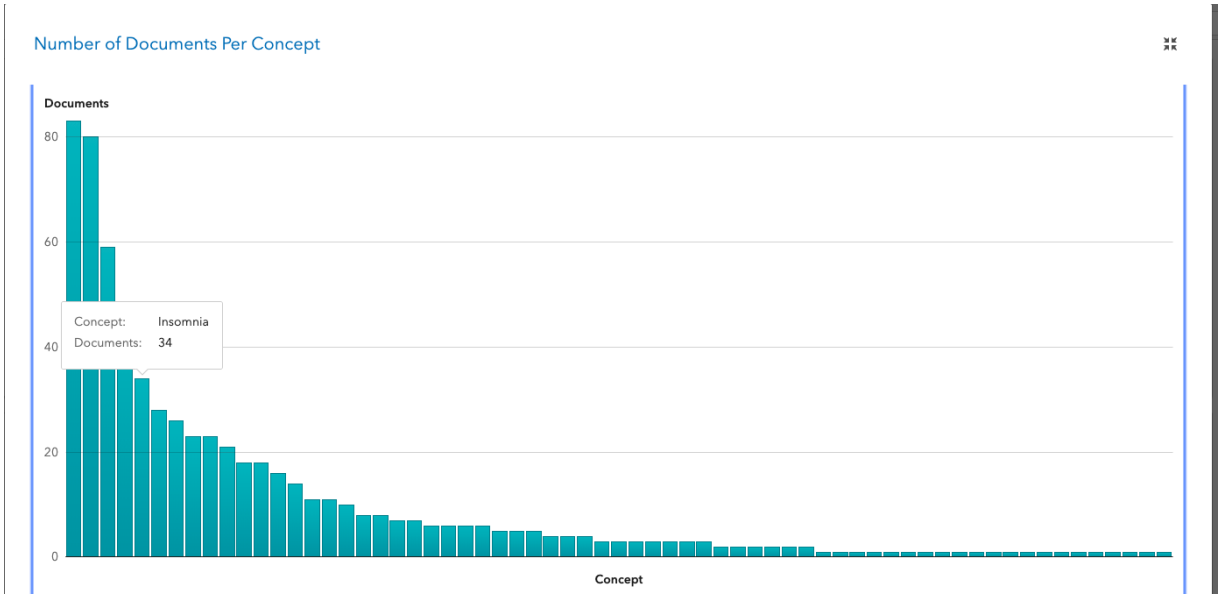
- **Scrolling or hovering over third highest Bar**



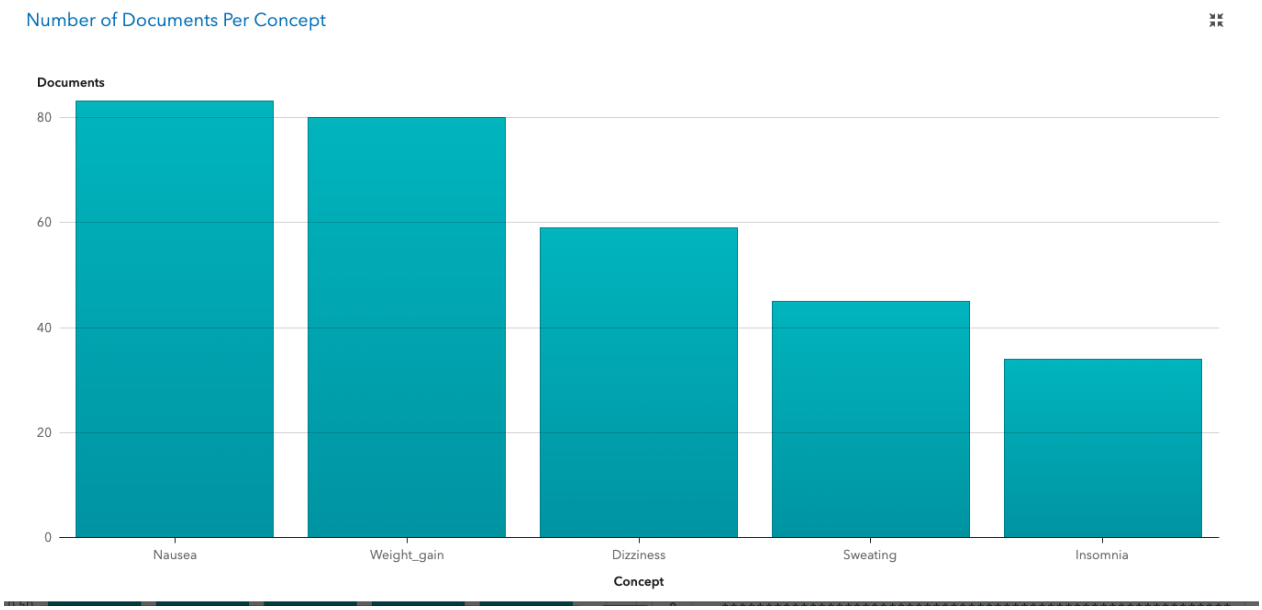
- **Scrolling or hovering over fourth highest Bar**



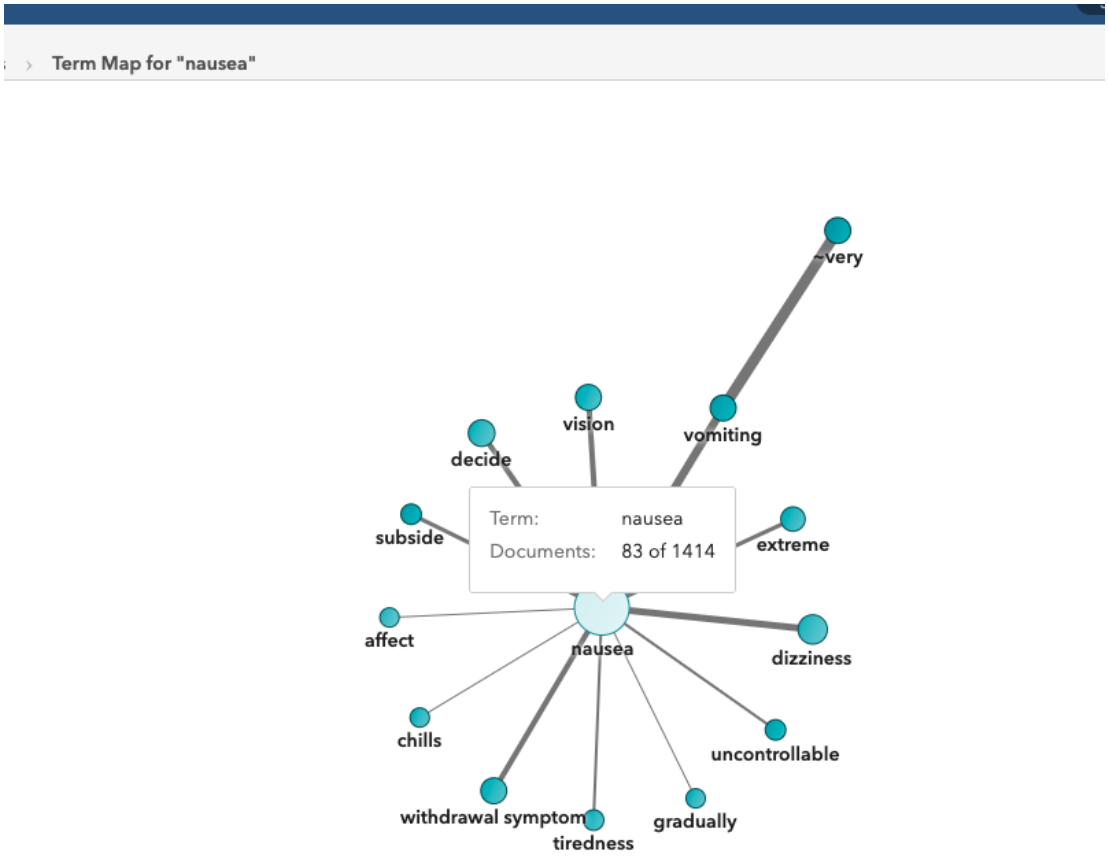
- **Scrolling or hovering over fifth highest Bar**



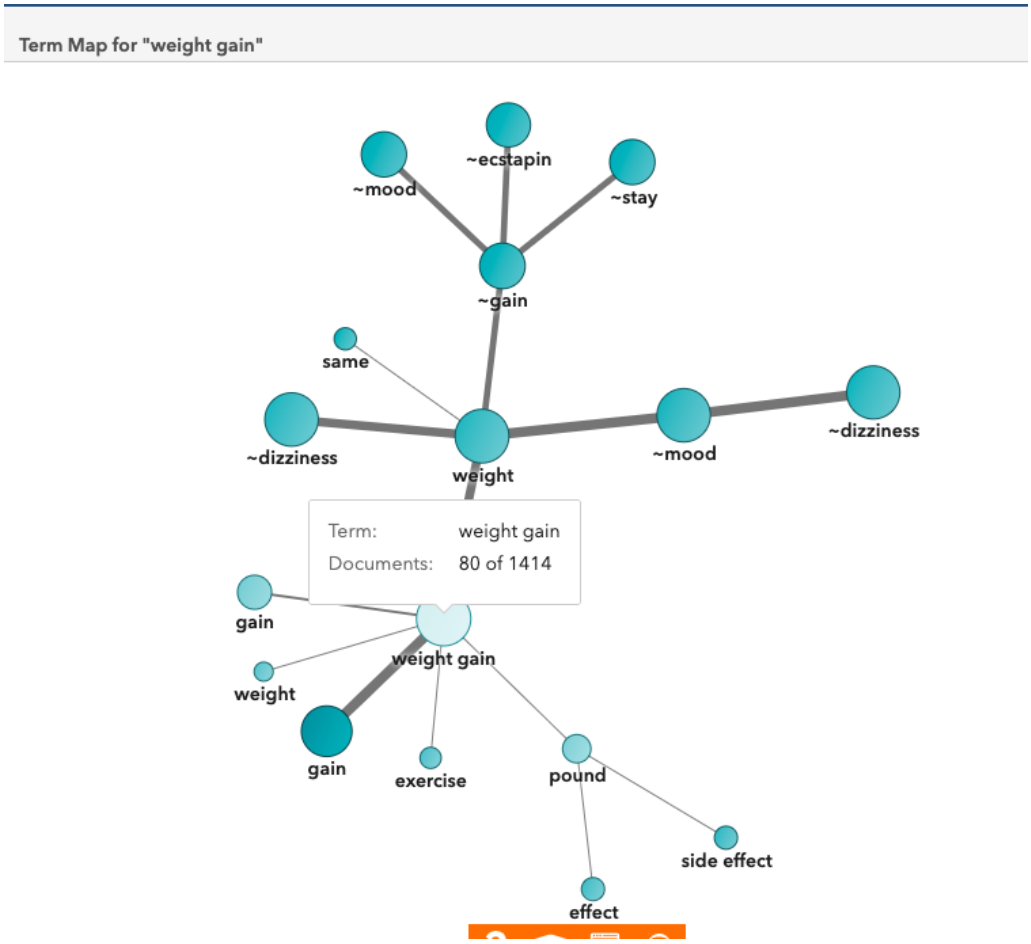
A separate Bar chart for those 5 most common concerns



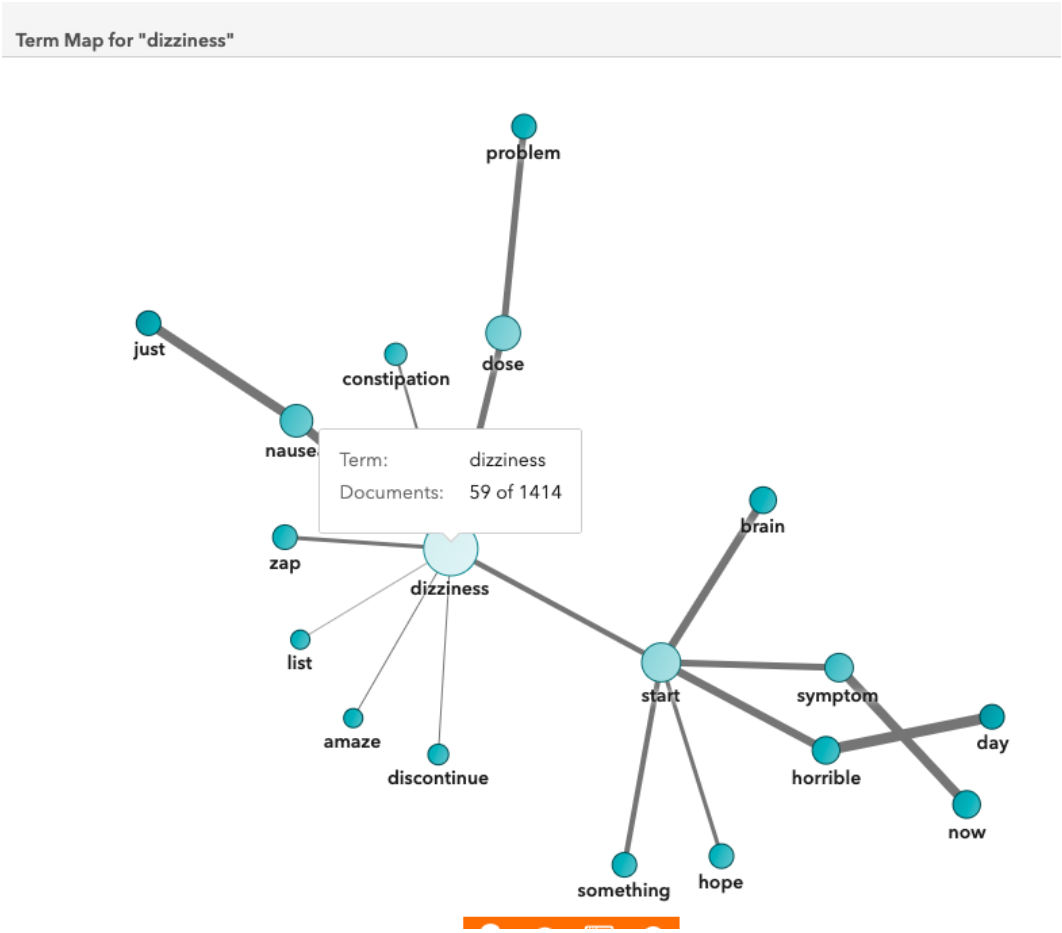
Term map for first highest common concern i.e. Nausea



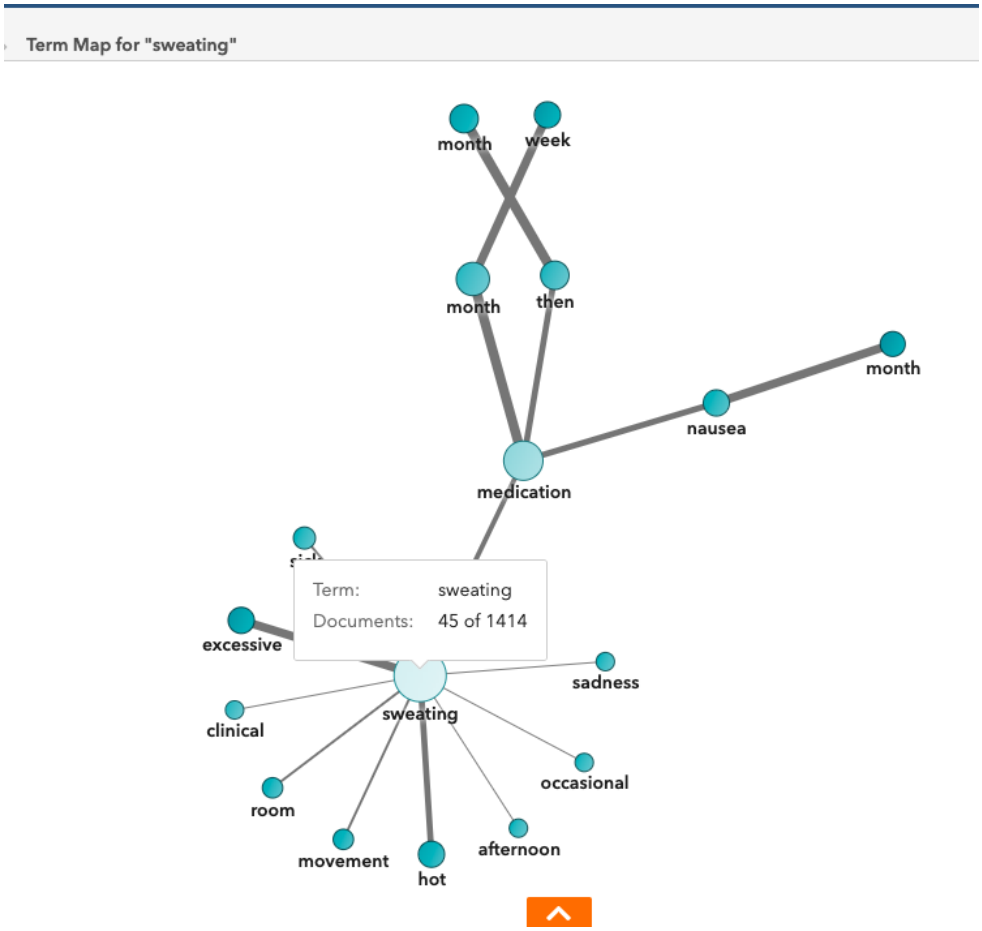
Term map for second highest common concern i.e. Weight gain



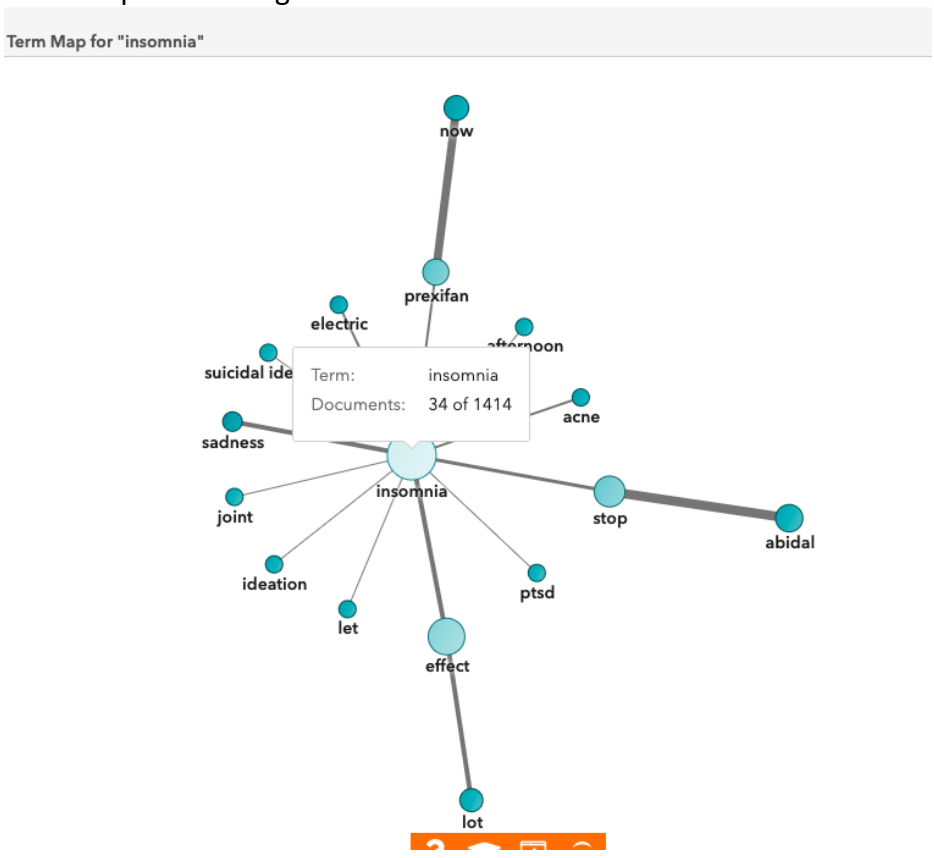
Term map for third highest common concern i.e. Dizziness



Term map for fourth highest common concern i.e. Sweating



Term map for fifth highest common concern i.e. Insomnia



Therefore, the ***first most common concern*** is ***Nausea*** with count ***83***, the ***second most common concern*** is ***Weight gain*** with count ***80***, the ***third most common concern*** is ***Dizziness*** with count ***59***, the ***fourth most common concern*** is ***Sweating*** with count ***45***, the ***fifth most common concern*** is ***Insomnia*** with count ***34***.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Task 5 (5 marks) - Provide useful additional information.

This final task is open. Provide additional charts that would help the analyst find additional useful information about drug effects on patients. For this task, in addition to explaining how the charts were produced and how to read them, it is important that you explain how the information is useful.

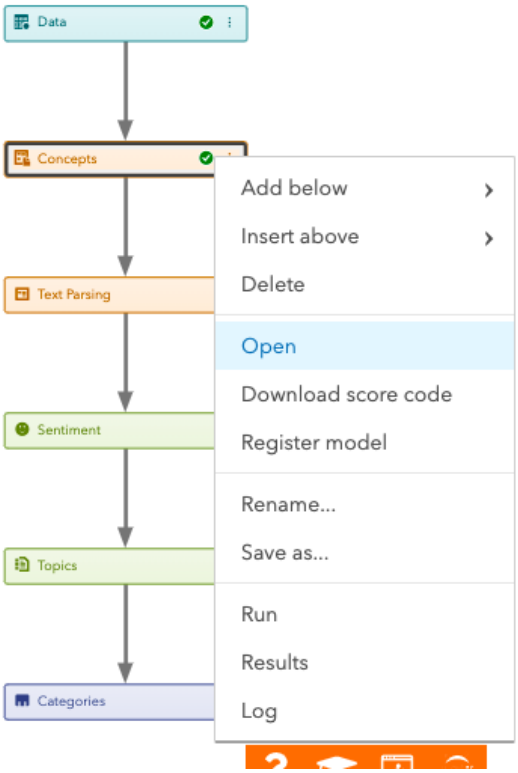
For Task 5, I used only SAS Viya for Learners.

Here, I will give a more brief information of Task 3. In the solution of Task 3, I found the medications that are associated with sleep issues i.e. sleep(Noun), sleep(Verb), sleep issue(Noun), sleep-aid(Noun), dont sleep(Noun), etc. But, there can be other classifiers also which indirectly mean sleep or any other sleeping issue. So, Dizziness and Insomnia are the two classifiers other than Sleeplessness, which are also related to sleeping issues.

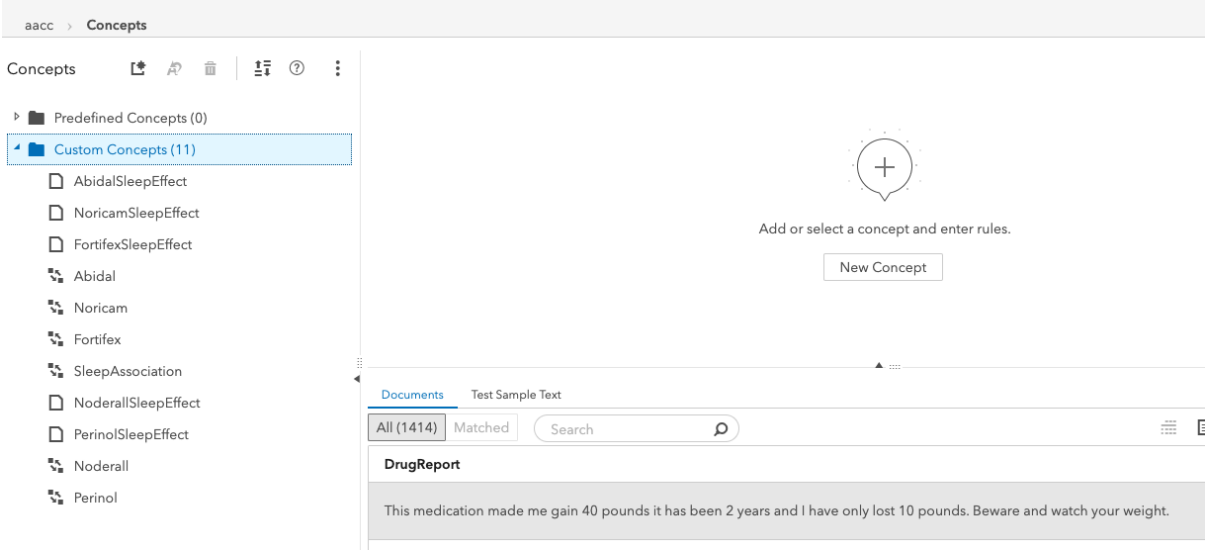
From Task 3, we got to know that ***medications*** that are ***associated with sleep issues*** are ***Fortifex***, ***Noricam, Noderall, Essequal, Abidal, and Perinol***.

Therefore, now I will also add Dizziness and Insomnia classifiers and call a concept from other.

First, I opened the Concepts Node.

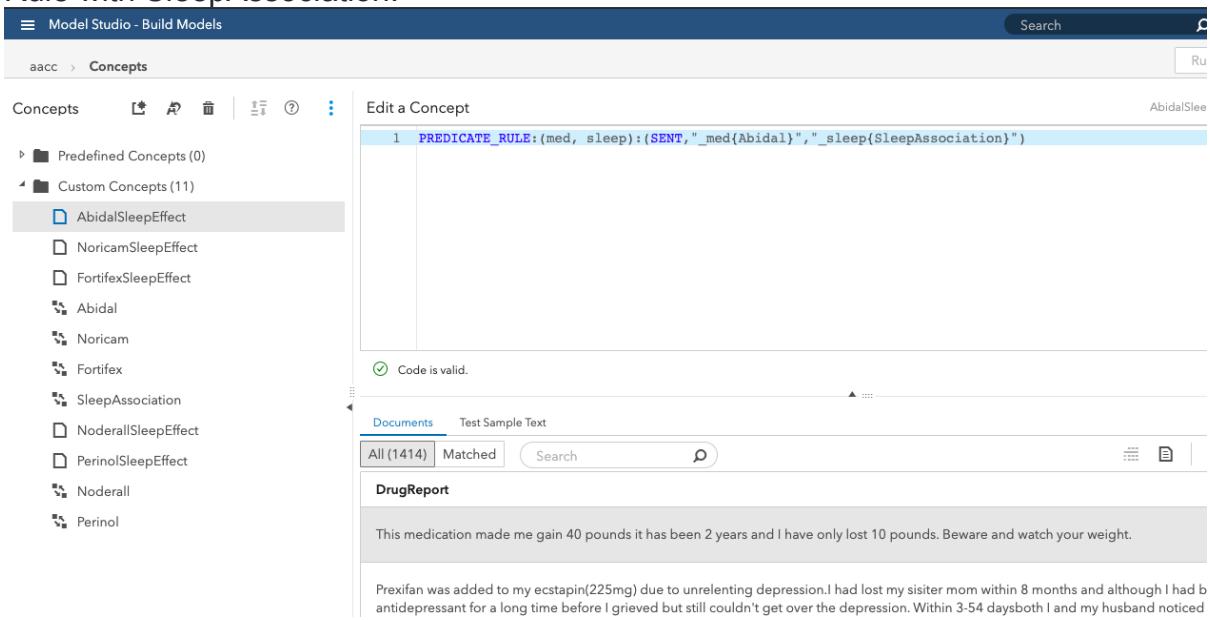


Then, made these Concepts.

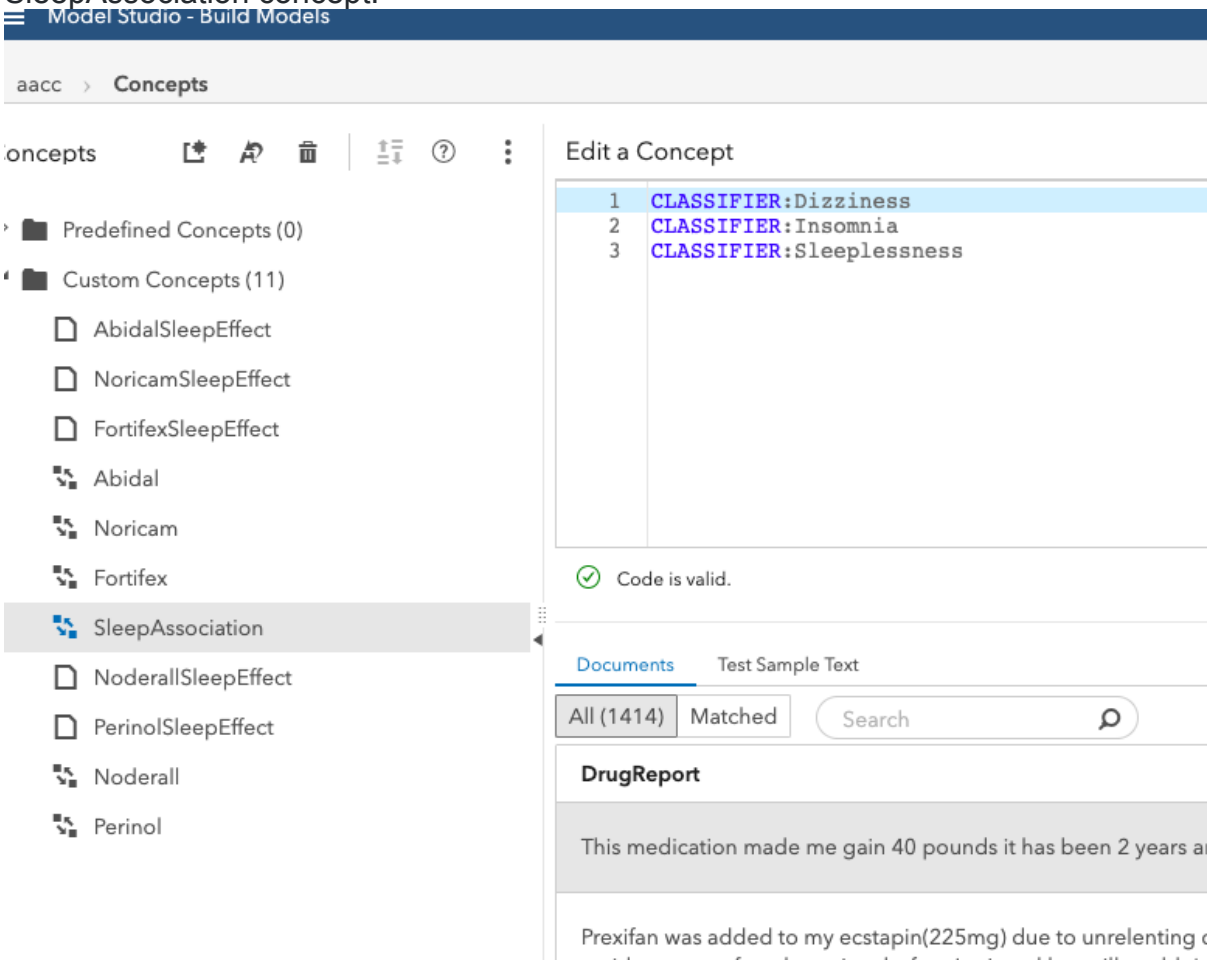


The 5 SleepEffect concepts are describing the predicate rule, which will find the association between the given medicine (name of the particular medicine that we will pass) and SleepAssociation (concept containing the classifiers of Dizziness, Insomnia, & Sleeplessness).

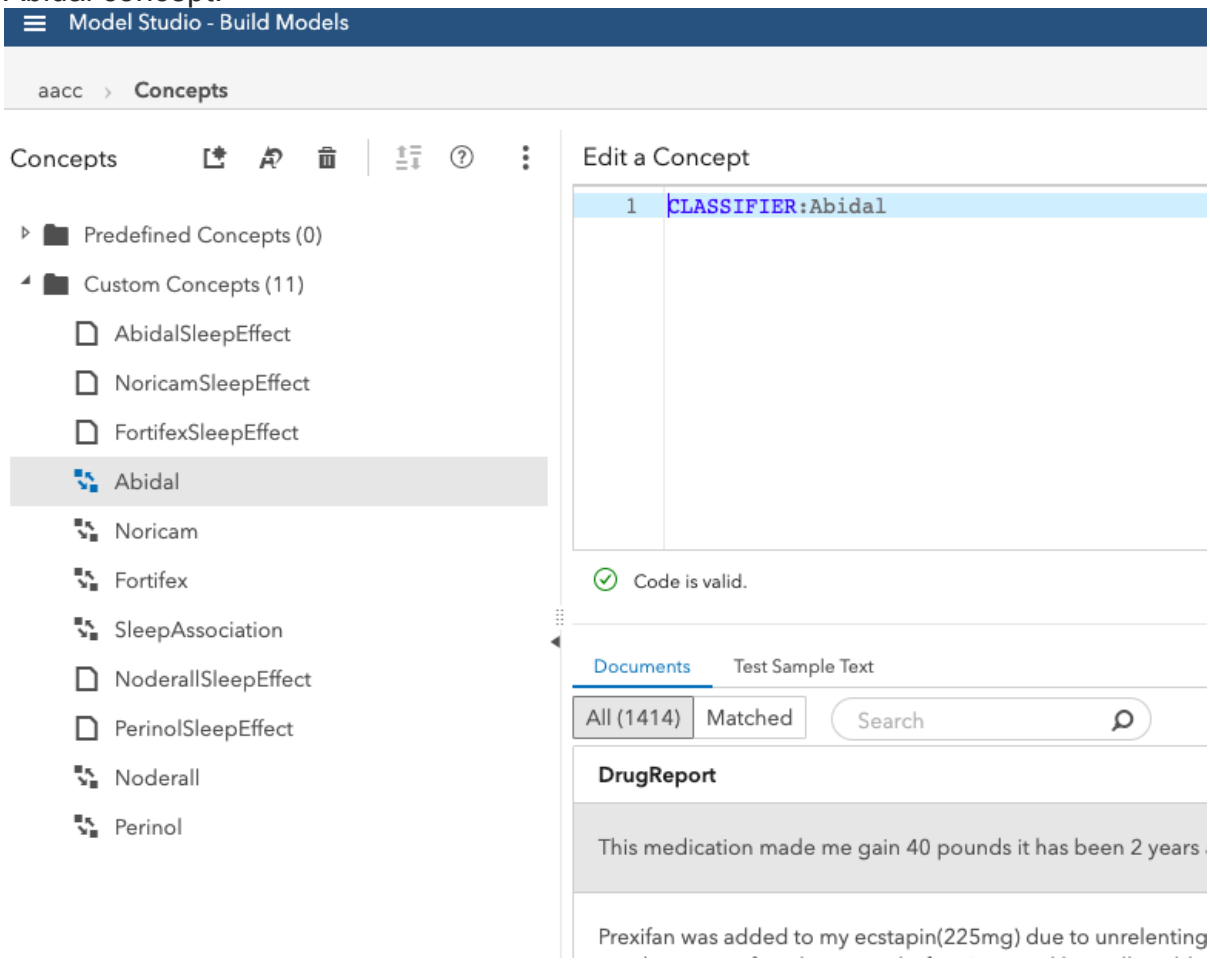
For example, as we can see below, Abidal medication is passed in the Predicate Rule with SleepAssociation.



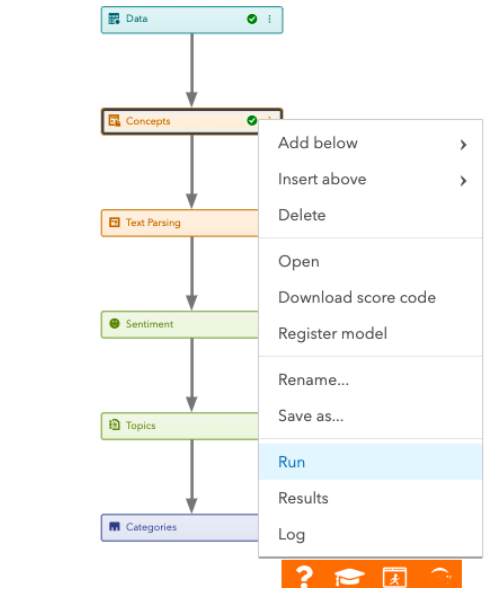
As we can see below, Dizziness, Insomnia, & Sleeplessness classifiers are passed in SleepAssociation concept.

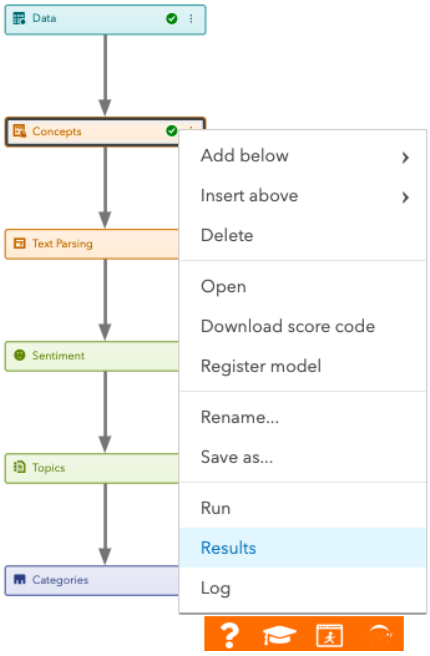


The 5 single medication named concepts are describing the classifiers for only that medication.
For example, as we can see below, Abidal medication's classifier is passed in the Abidal concept.

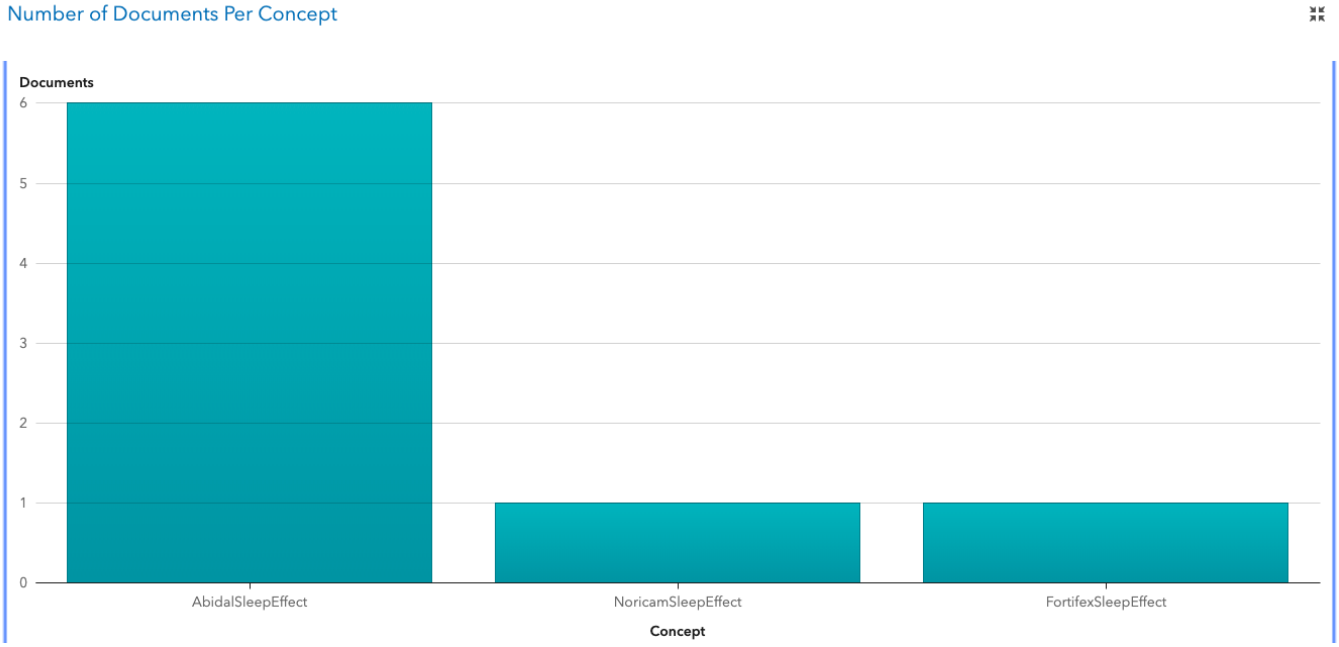


Now, we will Run and see the Results of this pipeline.

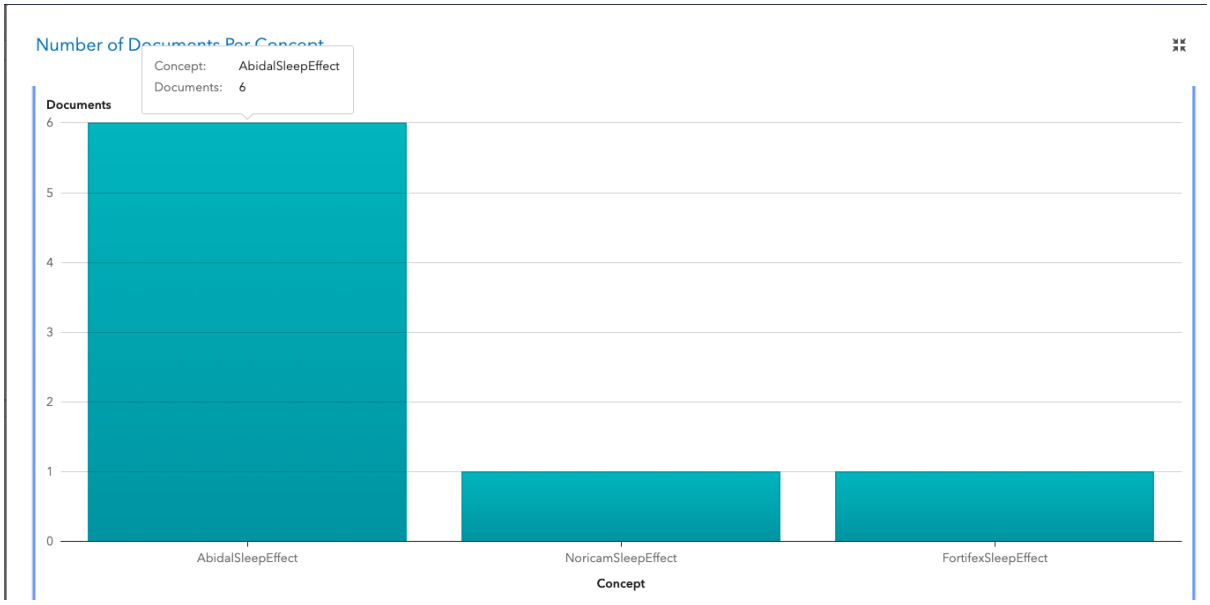




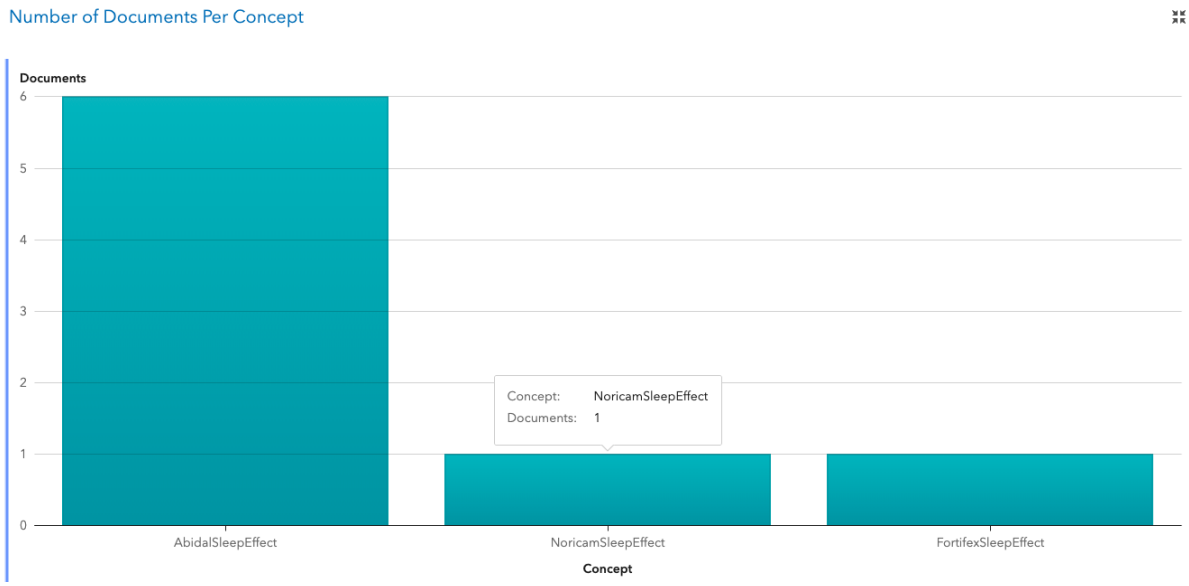
Finally, we will get the bar graph representing the count for AbidalSleepEffect, NoricamSleepEffect, & FortifexSleepEffect.
By scrolling or hovering over the different bars we can get the count of each element i.e. AbidalSleepEffect, NoricamSleepEffect, & FortifexSleepEffect.



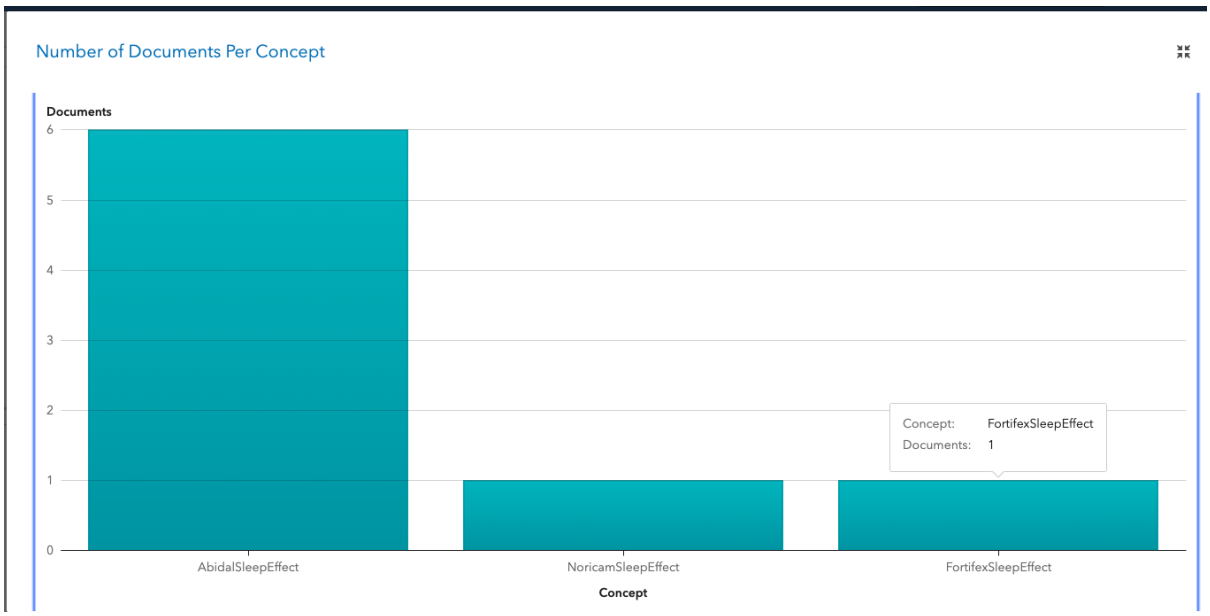
- **Scrolling or hovering over** AbidalSleepEffect



- **Scrolling or hovering over** NoricamSleepEffect



- **Scrolling or hovering over** FortifexSleepEffect



Therefore, AbidalSleepEffect has 6 count, NoricamSleepEffect has 1 count, & FortifexSleepEffect has 1 count. These counts are representing the association of a particular medication with Sleeplessness, Dizziness and Insomnia classifiers.