## **Attachments:**

1. Source Code: AmazonReviewAnalysis.zip
2. Data Set: Unzip AmazonReviewAnalysis,

go to AmazonReviewAnalysis/NegativeReviews.csv, AmazonReviewAnalysis/ positiveReviews.csv

1. Log.txt (for reference to see what happens when you run installation steps in this documents below)
2. ReadMe File

## **This project requires the following things in your machine to be able to work properly:**

## **Homebrew, gcc**

## **Python 3.5 & above**

## **Numpy, Scipy, Scikit-Learn**

## **NLTK (Natural Language Took Kit) & NLTK Data**

## **Eclipse**

## **PyDev Editor on Eclipse**

Please find below installation steps to setup your machine:

**STEP 1:** ruby -e "$(curl -fsSL <https://raw.githubusercontent.com/Homebrew/install/master/install>)"

**STEP 2:** brew install gcc

**STEP 3:** brew install python3

**STEP 4:** sudo pip3 install -U numpy

sudo pip3 install -U scipy

sudo pip3 install -U scikit-learn

**STEP 5:** sudo pip3 install -U nltk

*Note: Please refer log.txt for reference of installation steps between 1-4*

**STEP 6:** sudo python3 -m nltk.downloader

A screen as described below will appear, Please select first row with identifier as “**all**” and click **“download”.**

**Please note:** it may take some time to download entire data



You can verify as shown below if data is downloaded or not. You should see status as installed for these.





**STEP 7:** Please verify your java version (I used version 8) but if you have JDK installed you do not need to do any thing.

**Please go to Terminal and Execute:** java –version

**Output:**

java version "1.8.0\_111"

Java(TM) SE Runtime Environment (build 1.8.0\_111-b14)

Java HotSpot(TM) 64-Bit Server VM (build 25.111-b14, mixed mode)

You can also download java for the following websites:

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

<https://support.apple.com/kb/DL1572?locale=en_US>

**STEP 8:** Download Eclipse based on your java version compatibility. Minimum required is 1.6.

You can also download eclipse from

<http://www.eclipse.org/downloads/packages/eclipse-ide-java-ee-developers/keplerr>

**STEP 9:** Install PyDev in Eclipse, go to **Help -> Eclipse Marketplace**….



Search for **Pydev -> click install**



**Follow next 4 steps as shown below:**







Now reopen Eclipse, Go to **Window -> Perspective -> Open Perspective -> Other…**



You will find that python editor is installed in eclipse



Please find below instructions to setup and run the project on your machine:

**Import Project:**



Eclipse will prompt you a message “Python interpreter is not currently configured”. Please select **Advanced Auto-Config.**

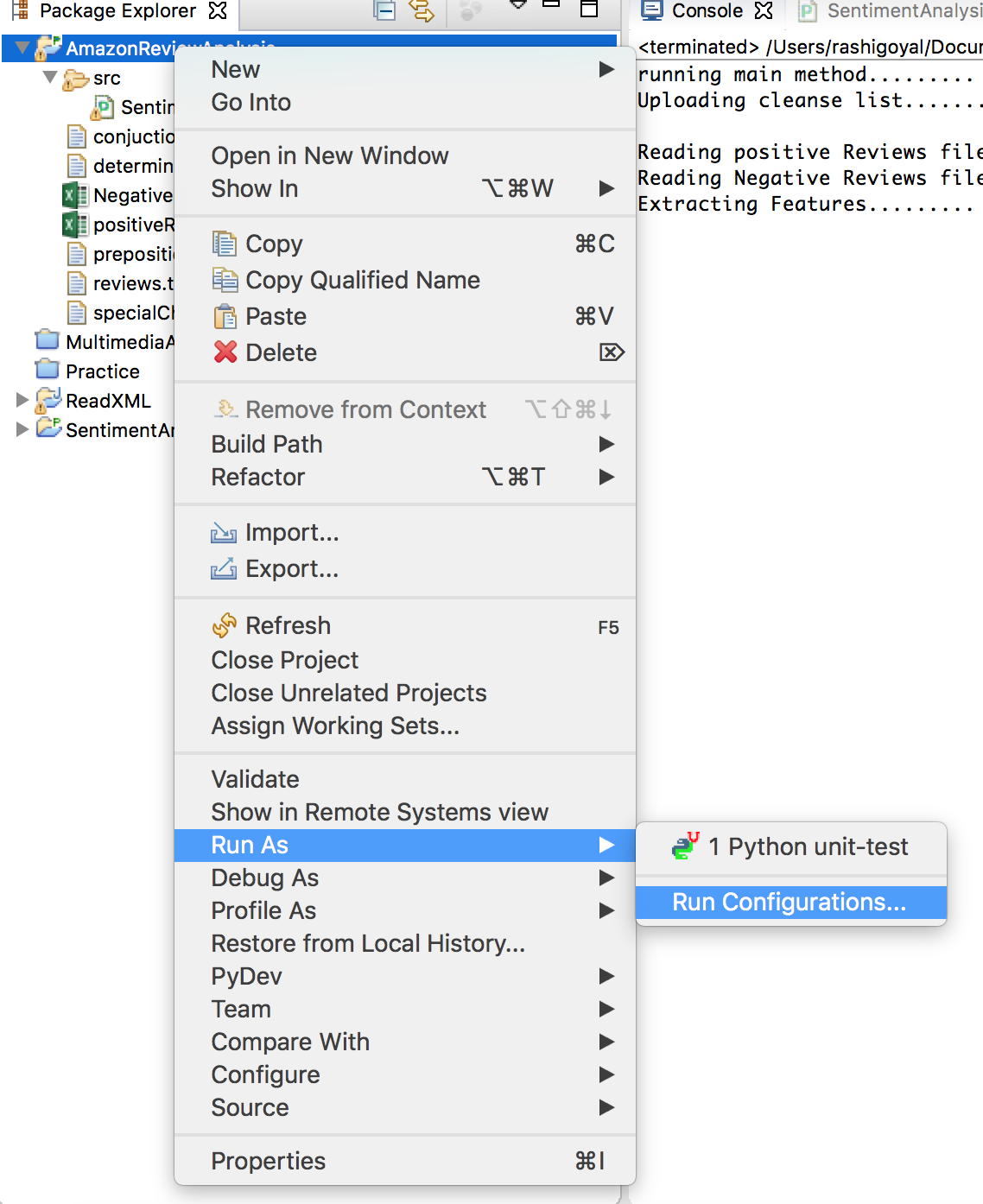


You can also go to **Eclipse -> preferences -> Pydev ->Interpreters -> python Interpreters** & Click on **Advanced Auto-Config.**

**Please select Python 3.5**



You should not see any issues with the project now. If you still see any issues with imports. Please right click project folder go to **Run as -> Run Configurations**

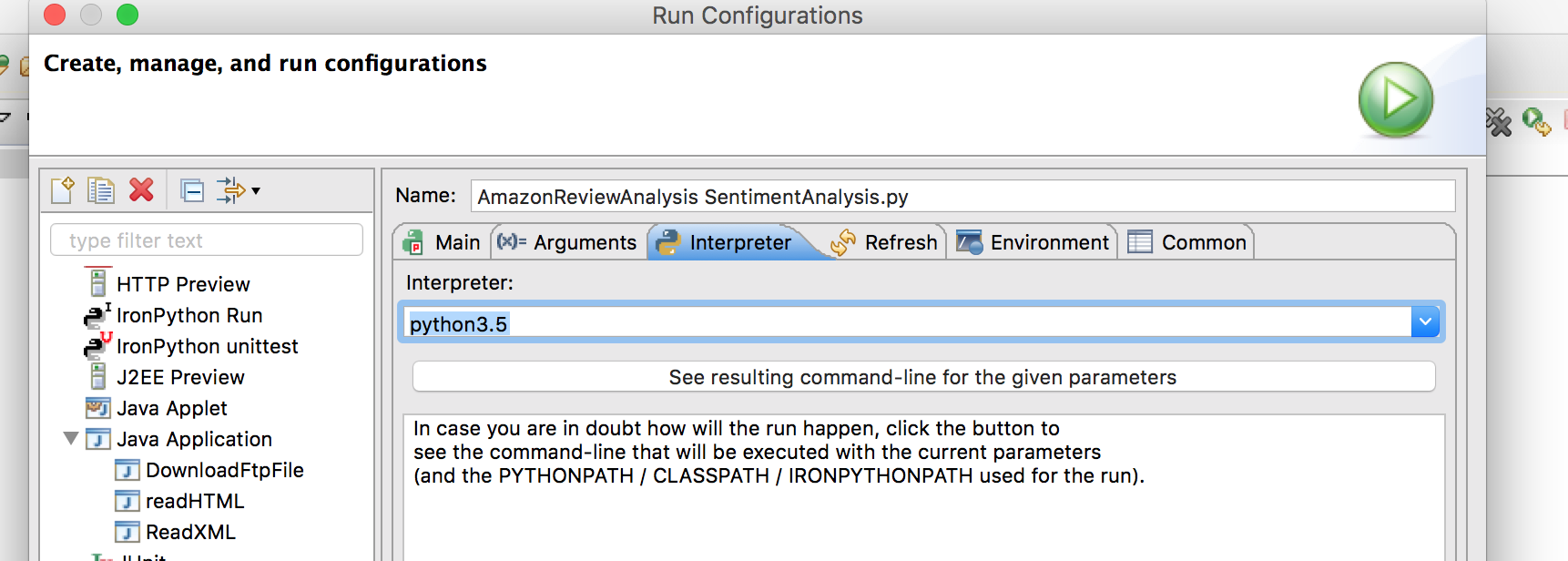


Go to interpreter and select Python 3.5 as your interpreter.

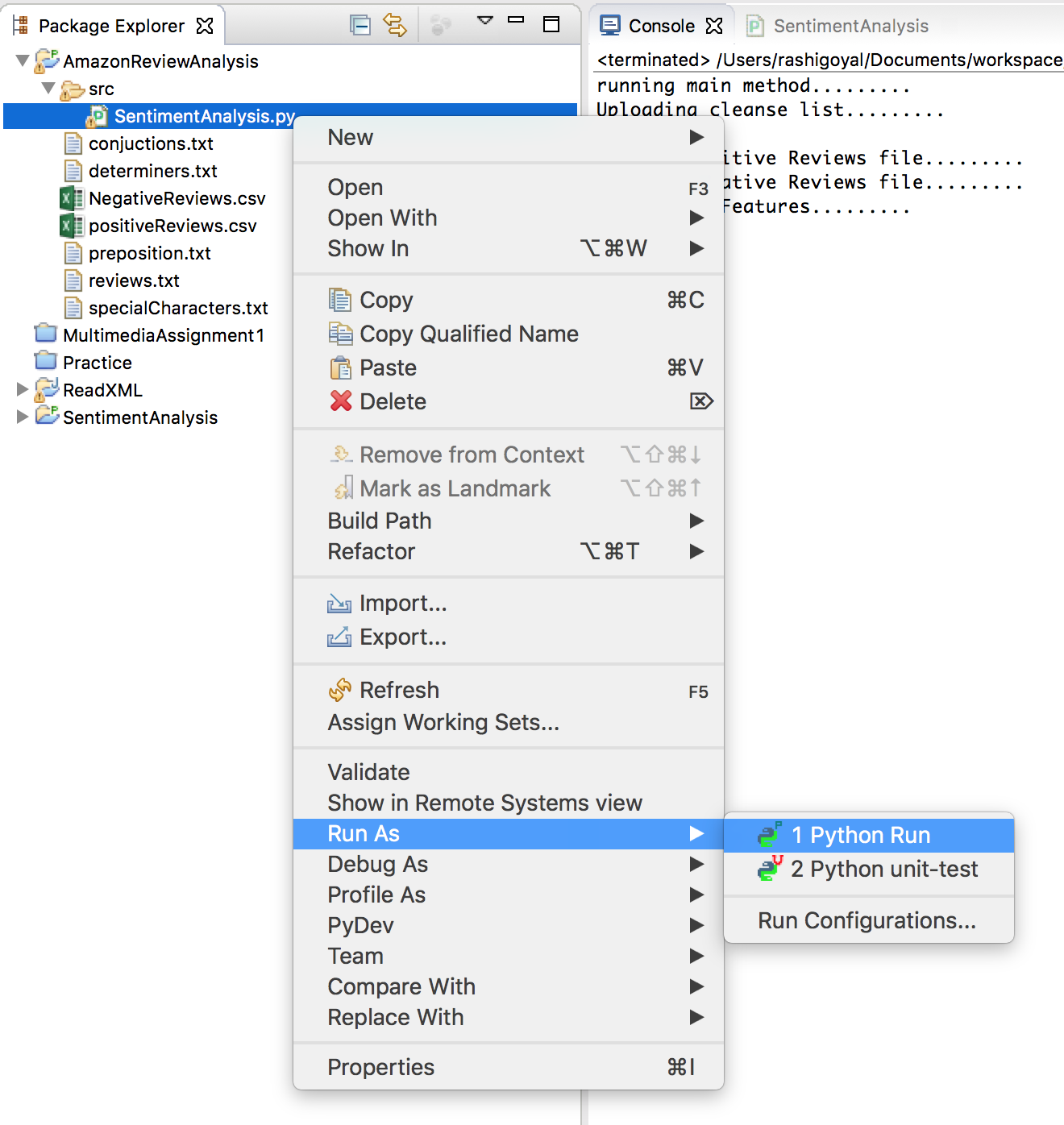
Please note: It may not show you python 3.5 as sometimes it is saved as a different alias name.

Please go to **Eclipse -> preferences -> Pydev ->Interpreters -> python Interpreters**

and check the alias name given for Python 3.5 interpreter

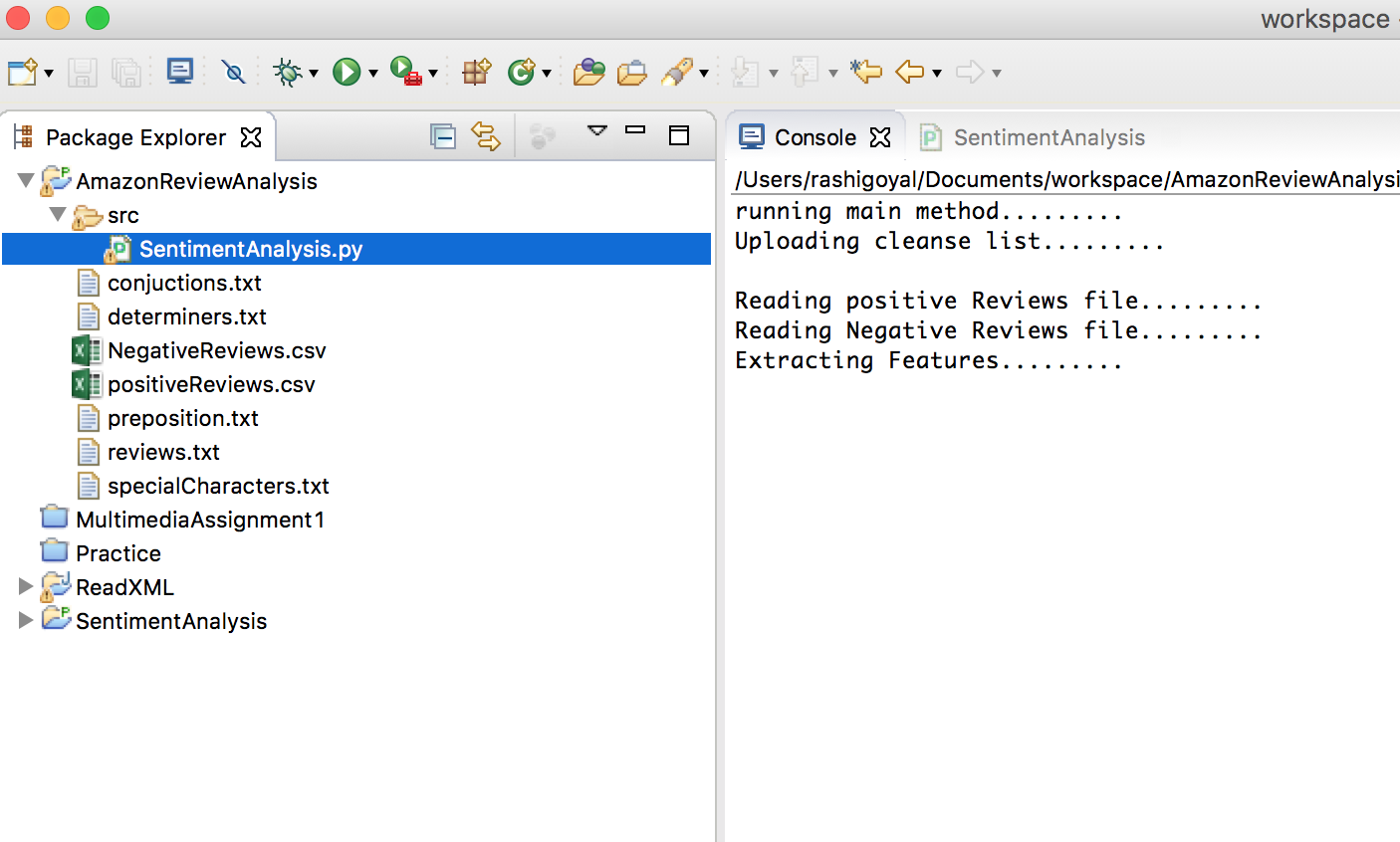


Go to project Explorer -> AmazonReviewAnalysis -> src -> (**Right click on**) SentimentAnalsis.py -> Runas -> Python Run



You should see the program is starting to execute in console Window.

**Please note:** It will take a while after **“Extracting Features…….”** As it is time consuming to process the data set.



**Program Output:**

running main method.........

Uploading cleanse list.........

Reading positive Reviews file.........

Reading Negative Reviews file.........

Extracting Features.........

Splitting Features into Training & Test Features.........

~~~~~~~~~~~~~~~ MultinomialNB Classifier ~~~~~~~~~~~~~~~

MultinomialNB Classifier Training Completed

MultinomialNB Classifier Test Results

Length of Training Features15000

Length of Test Features5002

Accuracy:0.8358656537385046

Positive precision: 0.840080971659919

Positive recall: 0.8296681327469012

Negative precision: 0.8317535545023697

Negative recall: 0.8420631747301079

~~~~~~~~~~~~~~~Classification report~~~~~~~~~~~~~~~

              precision    recall  f1-score   support

   Negative       0.83      0.84      0.84      2501

   Positive       0.84      0.83      0.83      2501

avg / total       0.84      0.84      0.84      5002

~~~~~~~~~~~~~~~Confusion matrix~~~~~~~~~~~~~~~

 [[2106  395]

 [ 426 2075]]

~~~~~~~~~~~~~~~ BernoulliNB Classifier ~~~~~~~~~~~~~~~

BernoulliNB Classifier Training Completed

BernoulliNB Classifier Test Results

Length of Training Features15000

Length of Test Features5002

Accuracy:0.7449020391843263

Positive precision: 0.7136379490756889

Positive recall: 0.8180727708916433

Negative precision: 0.7868852459016393

Negative recall: 0.6717313074770092

~~~~~~~~~~~~~~~Classification report~~~~~~~~~~~~~~~

              precision    recall  f1-score   support

   Negative       0.79      0.67      0.72      2501

   Positive       0.71      0.82      0.76      2501

avg / total       0.75      0.74      0.74      5002

~~~~~~~~~~~~~~~Confusion matrix~~~~~~~~~~~~~~~

 [[1680  821]

 [ 455 2046]]

~~~~~~~~~~~~~~~ Linear SVC Classifier ~~~~~~~~~~~~~~~

Linear SVC Classifier Training Completed

Linear SVC Classifier Test Results

Length of Training Features15000

Length of Test Features5002

Accuracy:0.8420631747301079

Positive precision: 0.8467774625050669

Positive recall: 0.835265893642543

Negative precision: 0.8374753451676529

Negative recall: 0.8488604558176729

~~~~~~~~~~~~~~~Classification report~~~~~~~~~~~~~~~

              precision    recall  f1-score   support

   Negative       0.84      0.85      0.84      2501

   Positive       0.85      0.84      0.84      2501

avg / total       0.84      0.84      0.84      5002

~~~~~~~~~~~~~~~Confusion matrix~~~~~~~~~~~~~~~

 [[2123  378]

 [ 412 2089]]

~~~~~~~~~~~~~~~ Logistic Regression Classifier ~~~~~~~~~~~~~~~

trainingDone

Length of Training Features15000

Length of Test Features5002

Accuracy:0.8630547780887645

Positive precision: 0.8679092382495948

Positive recall: 0.8564574170331867

Negative precision: 0.8583267561168114

Negative recall: 0.8696521391443423

~~~~~~~~~~~~~~~Classification report~~~~~~~~~~~~~~~

              precision    recall  f1-score   support

   Negative       0.86      0.87      0.86      2501

   Positive       0.87      0.86      0.86      2501

avg / total       0.86      0.86      0.86      5002

~~~~~~~~~~~~~~~Confusion matrix~~~~~~~~~~~~~~~

 [[2175  326]

 [ 359 2142]]