Team: 14  
Professor: Yugyung Lee

Name: Sneha Mishra  
Class ID: 19  
Email: [smccr@mail.umkc.edu](mailto:smccr@mail.umkc.edu)  
[MyGitHub](https://github.com/SnehaMishra28/Python-DeepLearning_Fall2018/wiki/Class-Project)

Technical Partner:  
Name: Raju Nekadi  
Class ID: 20  
Email: [rn8mh@mail.umkc.edu](mailto:rn8mh@mail.umkc.edu)  
[GitHub](https://github.com/rnekadi/CSEE5590_PYTHON_DEEPLEARNING_FALL2018)

**Introduction:**

In todays world most of the top institution around the world mostly depend upon the IT infrastructure availability, web site availability plays very important role.

In order to make sure their Web Infrastructure available 24/7 to their customer organization around the world are spending a lot money in various research and solutions.

The main proposal of this project address one such solution where we try to predict future traffics for website and helps plan for infrastructure accordingly.

**Objective:**

The main objective of this project is very simple analytical work of taking the historical data and predicting the future.

The objective of this project is to predict future web traffic for given web pages of their site that are popular and if there are any apparent trends, such as one specific page being viewed mostly by people in a particular country.

It mainly focuses on forecasting the traffic on multiple time series values.

**Motivation:**

Being in professional world we came across a situation for client where they were not ready to support large incoming web request and end up losing the business.

That is the main motive for this project to implement solution for one such situation.

**Technologies Used:**

* This project mainly involves use of Machine learning and Deep learning frameworks, models and tools.
* The first phase is to use Regression model from Machine Learning library and get our prediction.
* The second phase is to use Deep Learning model bases RNN (Recurrent neural network), LSTM, Keras.
* Facebook New Forecasting model prophet - fbprophet
* Matplotlib, Seaborn and Cross validation libraries.

**Approach:**

We divided the project in 2 parts: Model Analysis and Data Analysis, which will help us in the analysis and prediction.

* Model Analysis: Create a model (RNN, then LSTM), compile and fit the model, then calculate the loss and accuracy.
* Data Analysis: Get the different data forms from the available dataset, do the Feature Scale on the data, then analyze and visualize the different categories.

**Workflow:**

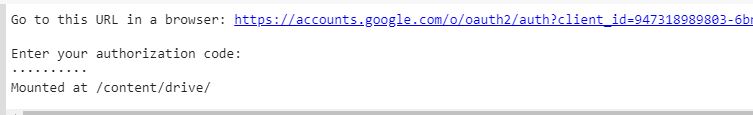
**Model Analysis:**

**Load the model:**

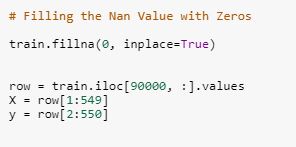
Load the model in csv format 

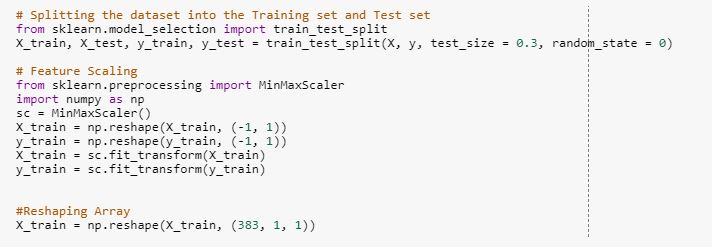
Display the first 5 rows of the dataset 

Or Mount the model using content drive in Colaboratory  
https://github.com/SnehaMishra28/Python-DeepLearning_Fall2018/raw/master/Project/Document/mount1.JPG



**Work on the dataset:**

Filling the Nan Value with Zeros.  
Our dataset had more than 150k rows, so we consider only 90k rows for our project.  
Also defining the X and Y for the model.  


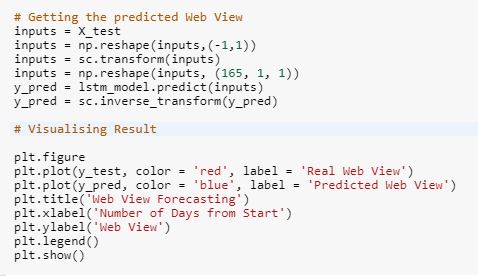
Splitting the dataset into the Training set and Test set.  
Feature Scaling.  
And Reshaping the Array.  


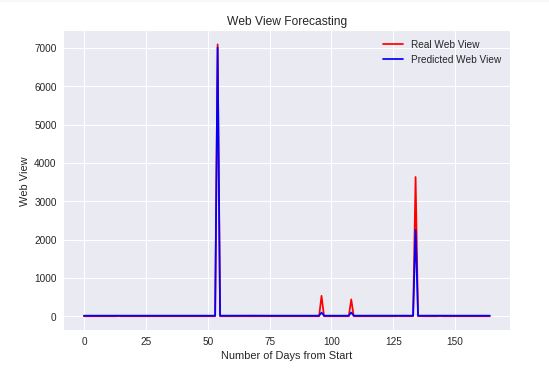
https://github.com/SnehaMishra28/Python-DeepLearning_Fall2018/raw/master/Project/Document/c3.JPG

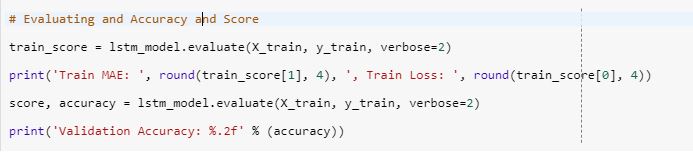
**Create the model:**

Importing the Keras libraries and packages for LSTM  
Initialising the RNN  
Adding the input layerand the LSTM layer  
Adding the output layer  
Compiling the RNN  
Fitting the LSTM to the Training set  


**Analyze and Predict on the Model:**

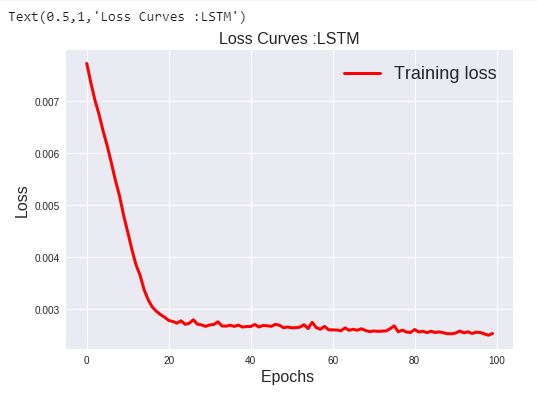
Getting the predicted Web View  
Visualizing Result  




Evaluating and Plotting the Accuracy and Score  


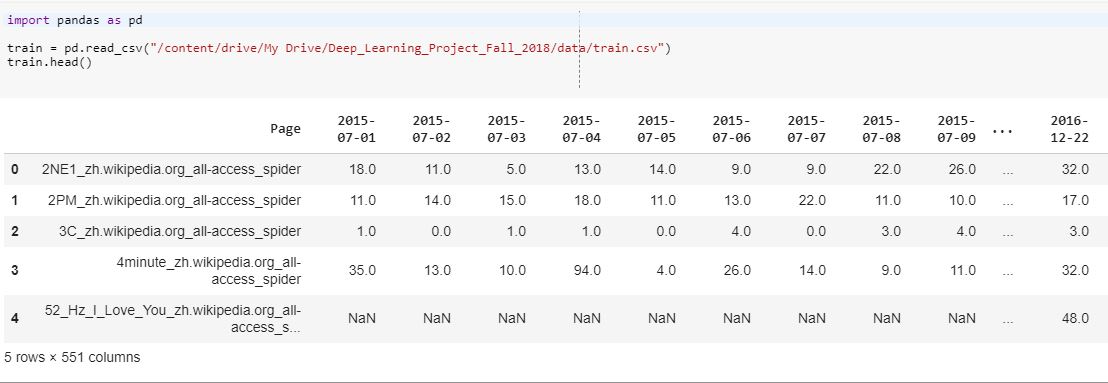
https://github.com/SnehaMishra28/Python-DeepLearning_Fall2018/raw/master/Project/Document/c8.JPG



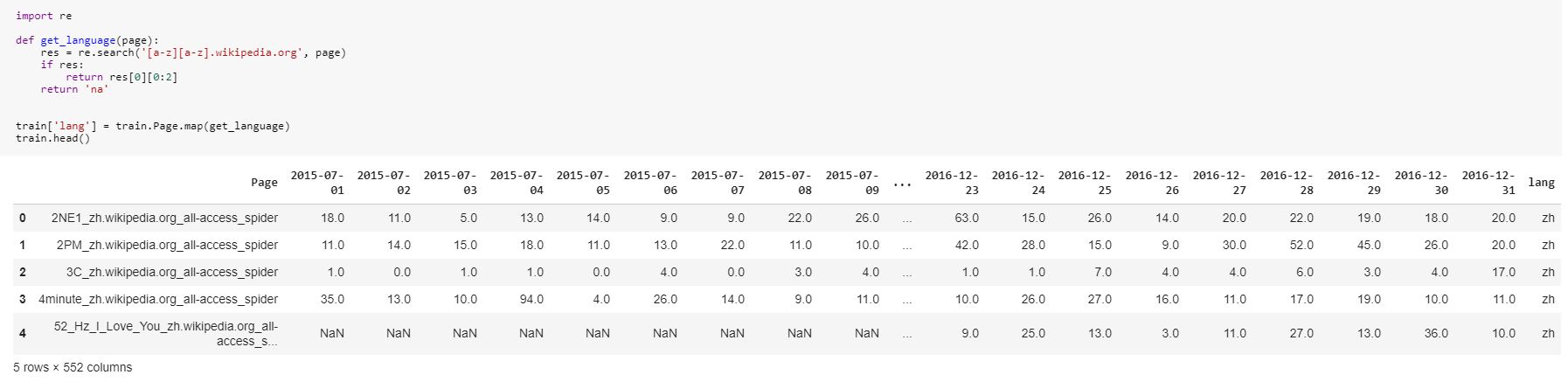


**Data Analysis:**

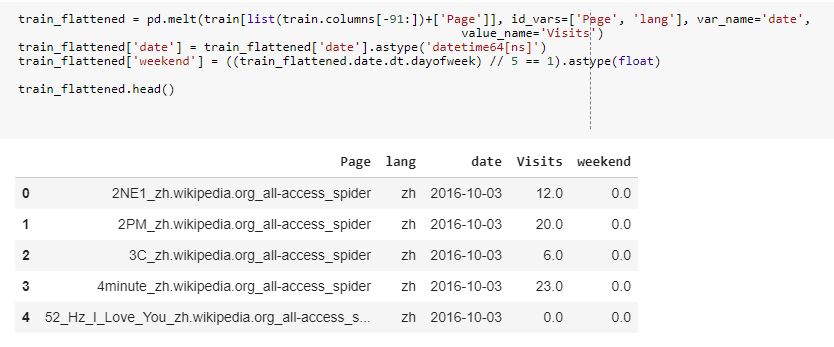
**Load the data and print the first 5 rows**



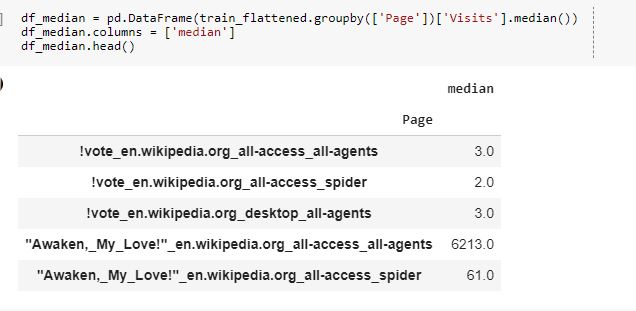
**Map the languages of each page and display it**



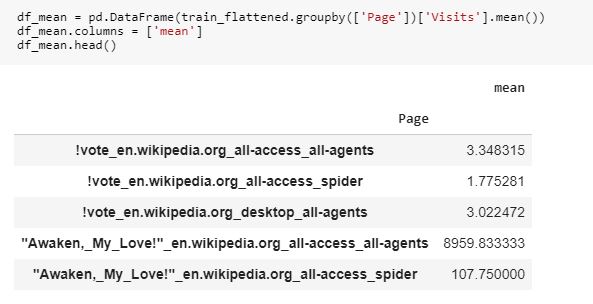
**Group the pages and print the language, visits and date**



**Get the median**



**Get the mean**

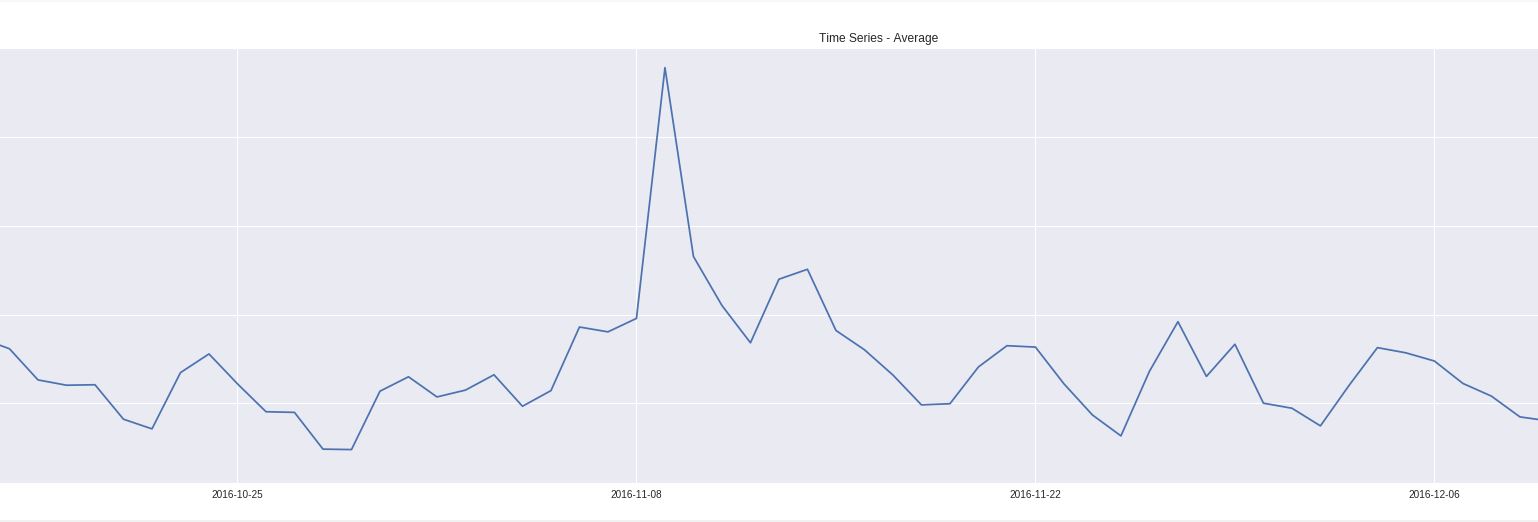


**Get the weekdays of the visited site, to visualize the same**



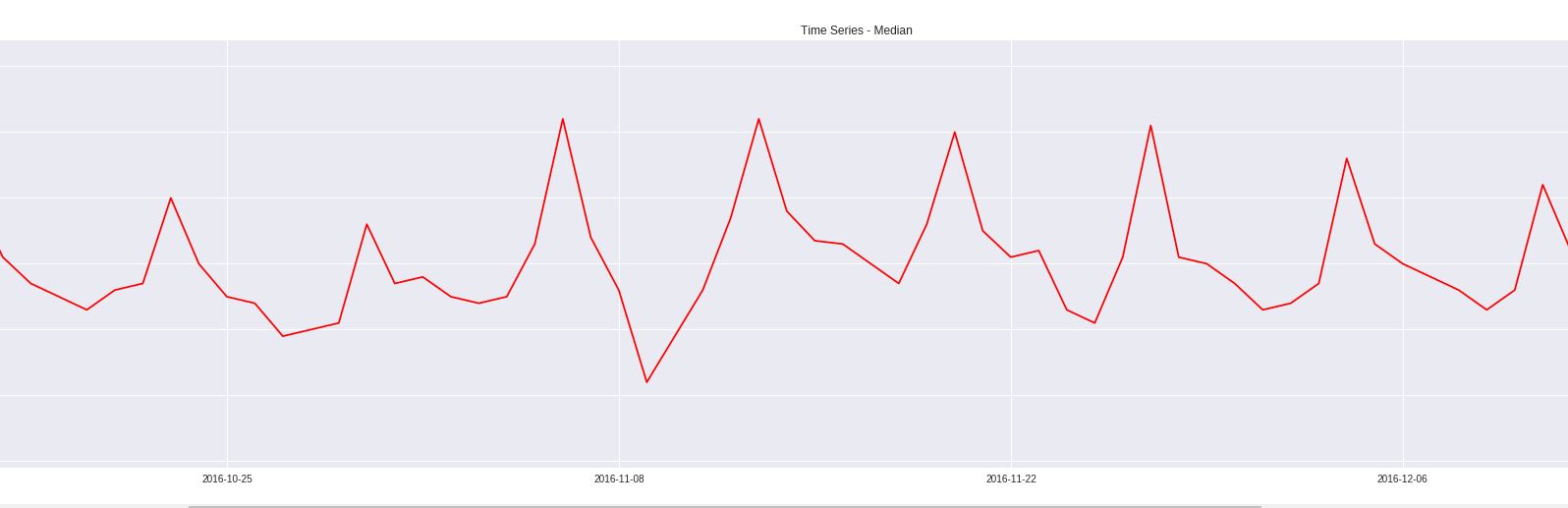
**View the mean pattern (need to be linear)**





**View the median pattern (need to be linear)**





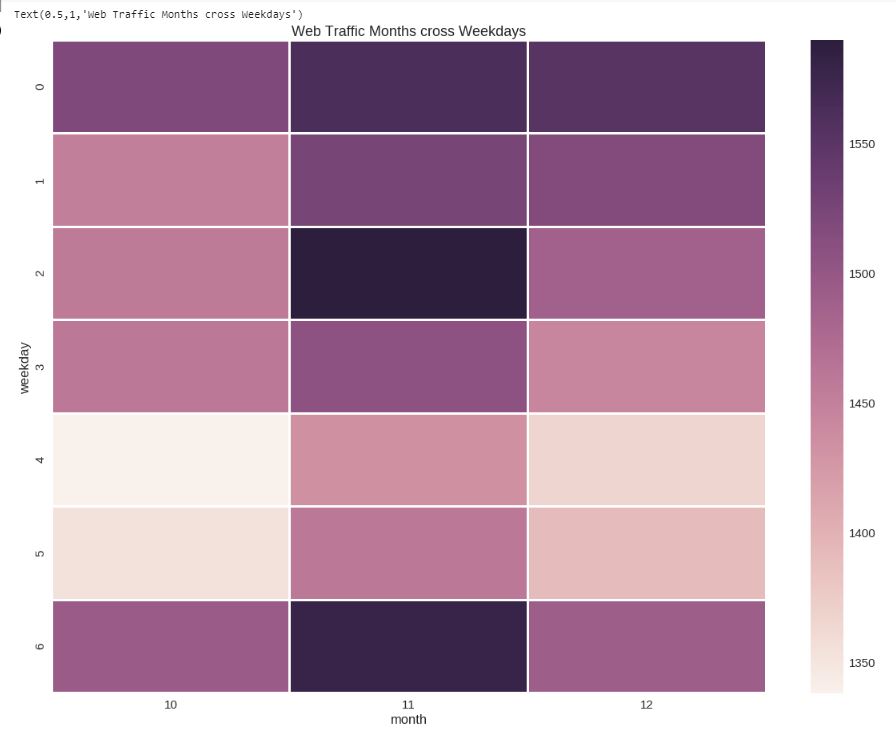
**View the standard pattern (need to be linear)**



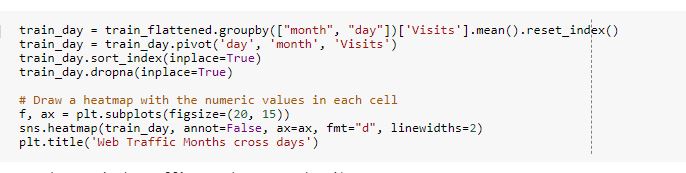


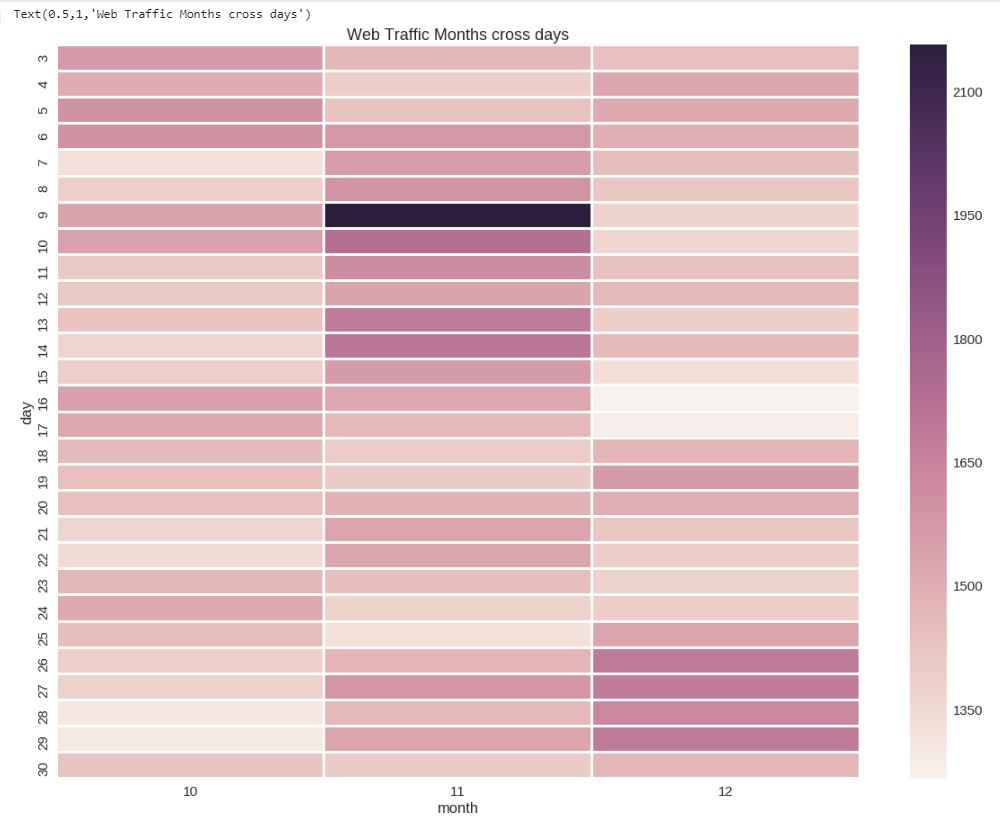
**Assigning the months and days, also view the Web Traffic Months cross Weekdays in heatmaps**



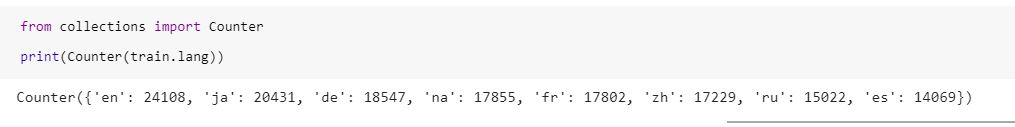


**View the Web Traffic Months cross days in heatmaps**



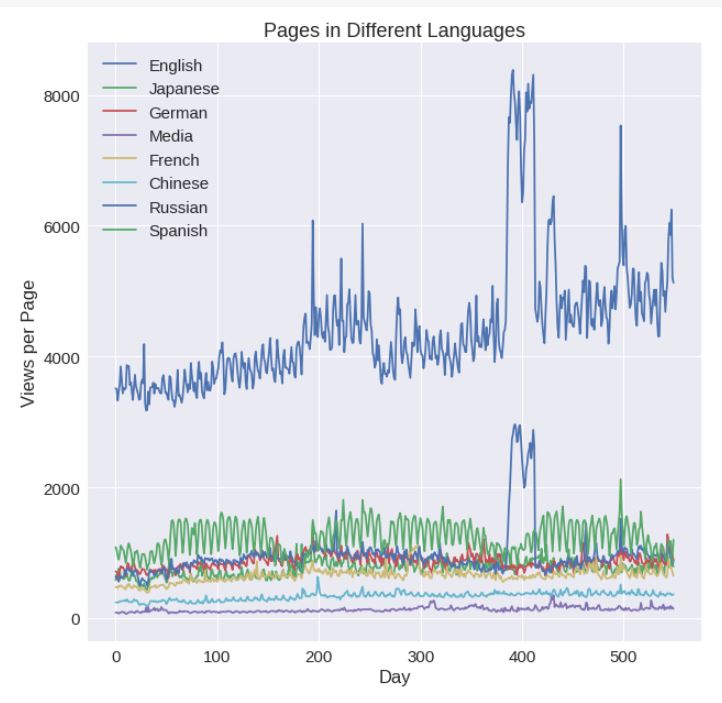


**Print the count of languages visited, then view it**



**View pages visited in different language**





**Dataset:**

This Web Traffic Time Series Forecasting (Forecast future traffic to Wikipedia pages). The file will look something like screenshot included below: 

**Parameters:**

* batch\_size = 10
* epochs = 100
* activation = 'relu'
* test\_size = 0.3
* row size = 90000

**Evaluation:**

Train MAE: 0.9373 , Train Loss: 0.0025  
Validation Accuracy: 0.94

**Conclusion:**

LSTM is better for time series model and gives best solution, compared to other models in our course.

**References:**

1. <https://en.wikipedia.org/wiki/Web_traffic>
2. <https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/>