In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import re
%matplotlib inline

train = pd.read_csv('train_1.csv').fillna(0)
train.head()
```

Out[1]:

	Page	2015-07-01	2015-07-02	2015-07-03	2015-07-04	2015-07-05	2015-07-06	2015-07-07	2015-07-08
0	2NE1_zh.wikipedia.org_all-access_spider	18.0	11.0	5.0	13.0	14.0	9.0	9.0	22.0
1	2PM_zh.wikipedia.org_all-access_spider	11.0	14.0	15.0	18.0	11.0	13.0	22.0	11.0
2	3C_zh.wikipedia.org_all-access_spider	1.0	0.0	1.0	1.0	0.0	4.0	0.0	3.0
3	4minute_zh.wikipedia.org_all-access_spider	35.0	13.0	10.0	94.0	4.0	26.0	14.0	9.0
4	52_Hz_I_Love_You_zh.wikipedia.org_all-access_s...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows × 551 columns

In [2]:

```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 145063 entries, 0 to 145062
Columns: 551 entries, Page to 2016-12-31
dtypes: float64(550), object(1)
memory usage: 609.8+ MB
```

In [3]:

```
def get_language(page):
    res = re.search('[a-z][a-z].wikipedia.org',page)
    if res:
        return res[0][0:2]
    return 'na'

train['lang'] = train.Page.map(get_language)

from collections import Counter

print(Counter(train.lang))
```

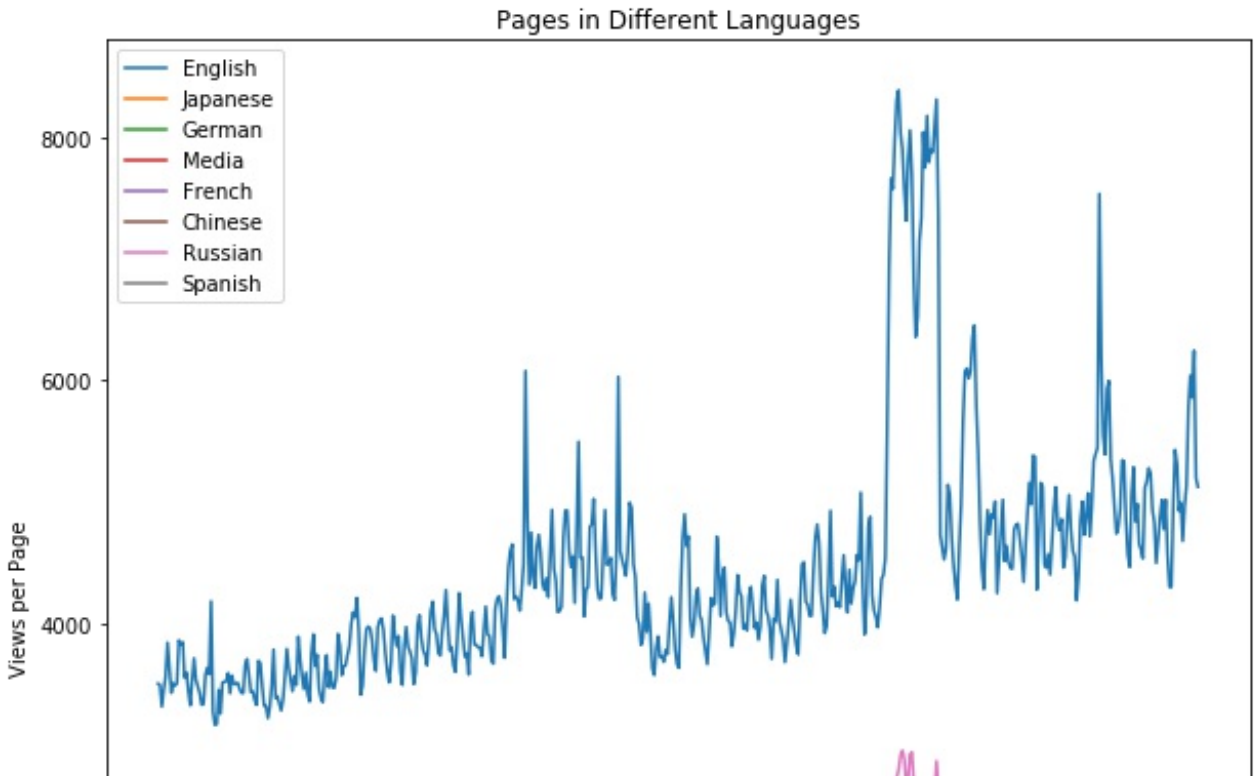
```
Counter({'en': 24108, 'ja': 20431, 'de': 18547, 'na': 17855, 'fr': 17802, '
zh': 17229, 'ru': 15022, 'es': 14069})
```

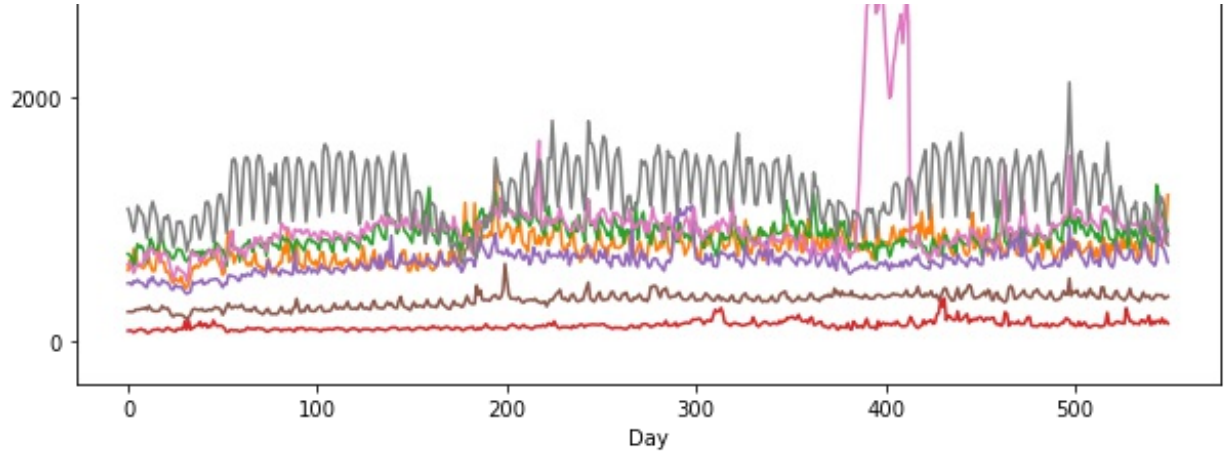
In [4]:

```
lang_sets = {}
lang_sets['en'] = train[train.lang=='en'].iloc[:,0:-1]
lang_sets['ja'] = train[train.lang=='ja'].iloc[:,0:-1]
lang_sets['de'] = train[train.lang=='de'].iloc[:,0:-1]
lang_sets['na'] = train[train.lang=='na'].iloc[:,0:-1]
lang_sets['fr'] = train[train.lang=='fr'].iloc[:,0:-1]
lang_sets['zh'] = train[train.lang=='zh'].iloc[:,0:-1]
lang_sets['ru'] = train[train.lang=='ru'].iloc[:,0:-1]
lang_sets['es'] = train[train.lang=='es'].iloc[:,0:-1]

sums = {}
for key in lang_sets:
    sums[key] = lang_sets[key].iloc[:,1:].sum(axis=0) / lang_sets[key].shape
[0]
```

In [5]:

```
days = [r for r in range(sums['en'].shape[0])]

fig = plt.figure(1,figsize=[10,10])
plt.ylabel('Views per Page')
plt.xlabel('Day')
plt.title('Pages in Different Languages')
labels={'en':'English','ja':'Japanese','de':'German',
        'na':'Media','fr':'French','zh':'Chinese',
        'ru':'Russian','es':'Spanish'
       }

for key in sums:
    plt.plot(days,sums[key],label = labels[key] )

plt.legend()
plt.show()
```





In [7]:

```
# For each language get highest few pages
npages = 5
top_pages = {}
for key in lang_sets:
    print(key)
    sum_set = pd.DataFrame(lang_sets[key][['Page']])
    sum_set['total'] = lang_sets[key].sum(axis=1)
    sum_set = sum_set.sort_values('total',ascending=False)
    print(sum_set.head(10))
    top_pages[key] = sum_set.index[0]
    print('\n\n')
```

```
en
                                                     Page         total
38573   Main_Page_en.wikipedia.org_all-access_all-agents  1.206618e+10
9774       Main_Page_en.wikipedia.org_desktop_all-agents  8.774497e+09
74114   Main_Page_en.wikipedia.org_mobile-web_all-agents  3.153985e+09
39180  Special:Search_en.wikipedia.org_all-access_all...  1.304079e+09
10403  Special:Search_en.wikipedia.org_desktop_all-ag...  1.011848e+09
74690  Special:Search_en.wikipedia.org_mobile-web_all...  2.921628e+08
39172  Special:Book_en.wikipedia.org_all-access_all-a...  1.339931e+08
10399   Special:Book_en.wikipedia.org_desktop_all-agents  1.332859e+08
33644       Main_Page_en.wikipedia.org_all-access_spider  1.290204e+08
34257  Special:Search_en.wikipedia.org_all-access_spider  1.243102e+08



ja
                                                     Page        total
120336      メインページ_ja.wikipedia.org_all-access_all-agents  210753795.0
86431          メインページ_ja.wikipedia.org_desktop_all-agents  134147415.0
123025       特別:検索_ja.wikipedia.org_all-access_all-agents   70316929.0
89202           特別:検索_ja.wikipedia.org_desktop_all-agents   69215206.0
57309       メインページ_ja.wikipedia.org_mobile-web_all-agents   66459122.0
119609    特別:最近の更新_ja.wikipedia.org_all-access_all-agents   17662791.0
88897        特別:最近の更新_ja.wikipedia.org_desktop_all-agents   17627621.0
119625        真田信繁_ja.wikipedia.org_all-access_all-agents   10793039.0
123292  特別:外部リンク検索_ja.wikipedia.org_all-access_all-agents
10331191.0
89463      特別:外部リンク検索_ja.wikipedia.org_desktop_all-agents
10327917.0
```

de

	Page	total
139119	Wikipedia:Hauptseite_de.wikipedia.org_all-acce...	1.603934e+09
116196	Wikipedia:Hauptseite_de.wikipedia.org_mobile-w...	1.112689e+09
67049	Wikipedia:Hauptseite_de.wikipedia.org_desktop_...	4.269924e+08
140151	Spezial:Suche_de.wikipedia.org_all-access_all-...	2.234259e+08
66736	Spezial:Suche_de.wikipedia.org_desktop_all-agents	2.196368e+08
140147	Spezial:Anmelden_de.wikipedia.org_all-access_a...	4.029181e+07
138800	Special:Search_de.wikipedia.org_all-access_all...	3.988154e+07
68104	Spezial:Anmelden_de.wikipedia.org_desktop_all-...	3.535523e+07
68511	Special:MyPage/toolserverhelferleinconfig.js_d...	3.258496e+07
137765	Hauptseite_de.wikipedia.org_all-access_all-agents	3.173246e+07

na

	Page	total
45071	Special:Search_commons.wikimedia.org_all-acces...	67150638.0
81665	Special:Search_commons.wikimedia.org_desktop_a...	63349756.0
45056	Special:CreateAccount_commons.wikimedia.org_al...	53795386.0
45028	Main_Page_commons.wikimedia.org_all-access_all...	52732292.0
81644	Special:CreateAccount_commons.wikimedia.org_de...	48061029.0
81610	Main_Page_commons.wikimedia.org_desktop_all-ag...	39160923.0
46078	Special:RecentChangesLinked_commons.wikimedia....	28306336.0
45078	Special:UploadWizard_commons.wikimedia.org_all...	23733805.0
81671	Special:UploadWizard_commons.wikimedia.org_des...	22008544.0
82680	Special:RecentChangesLinked_commons.wikimedia....	21915202.0

fr

	Page	total
27330	Wikipédia:Accueil_principal_fr.wikipedia.org_a...	868480667.0
55104	Wikipédia:Accueil_principal_fr.wikipedia.org_m...	611302821.0
7344	Wikipédia:Accueil_principal_fr.wikipedia.org_d...	239589012.0
27825	Spécial:Recherche_fr.wikipedia.org_all-access_...	95666374.0
8221	Spécial:Recherche_fr.wikipedia.org_desktop_all...	88448938.0
26500	Sp?cial:Search_fr.wikipedia.org_all-access_all...	76194568.0
6978	Sp?cial:Search_fr.wikipedia.org_desktop_all-ag...	76185450.0
131296	Wikipédia:Accueil_principal_fr.wikipedia.org_a...	63860799.0
26993	Organisme_de_placement_collectif_en_valeurs_mo...	36647929.0
7213	Organisme_de_placement_collectif_en_valeurs_mo...	36624145.0

zh

	Page	total
28727	Wikipedia:首页_zh.wikipedia.org_all-access_all-a...	123694312.0
61350	Wikipedia:首页_zh.wikipedia.org_desktop_all-agents	66435641.0
105844	Wikipedia:首页_zh.wikipedia.org_mobile-web_all-a...	50887429.0
28728	Special:搜索_zh.wikipedia.org_all-access_all-agents	48678124.0
61351	Special:搜索_zh.wikipedia.org_desktop_all-agents	48203843.0
28089	Running_Man_zh.wikipedia.org_all-access_all-ag...	11485845.0
30960	Special:链接搜索_zh.wikipedia.org_all-access_all-a...	10320403.0
63510	Special:链接搜索_zh.wikipedia.org_desktop_all-agents	10320336.0
60711	Running_Man_zh.wikipedia.org_desktop_all-agents	7968443.0
30446	瑯琊榜_(電視劇)_zh.wikipedia.org_all-access_all-agents	5891589.0

ru

ru

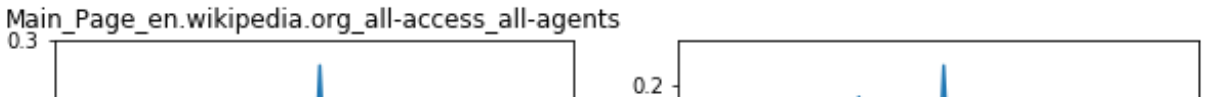
	Page	total
99322	Заглавная_страница_ru.wikipedia.org_all-access...	1.086019e+09
103123	Заглавная_страница_ru.wikipedia.org_desktop_al...	7.428800e+08
17670	Заглавная_страница_ru.wikipedia.org_mobile-web...	3.279304e+08
99537	Служебная:Поиск_ru.wikipedia.org_all-access_al...	1.037643e+08
103349	Служебная:Поиск_ru.wikipedia.org_desktop_all-a...	9.866417e+07
100414	Служебная:Ссылки_сюда_ru.wikipedia.org_all-acc...	2.510200e+07
104195	Служебная:Ссылки_сюда_ru.wikipedia.org_desktop...	2.505816e+07
97670	Special:Search_ru.wikipedia.org_all-access_all...	2.437457e+07
101457	Special:Search_ru.wikipedia.org_desktop_all-ag...	2.195847e+07
98301	Служебная:Вход_ru.wikipedia.org_all-access_all...	1.216259e+07

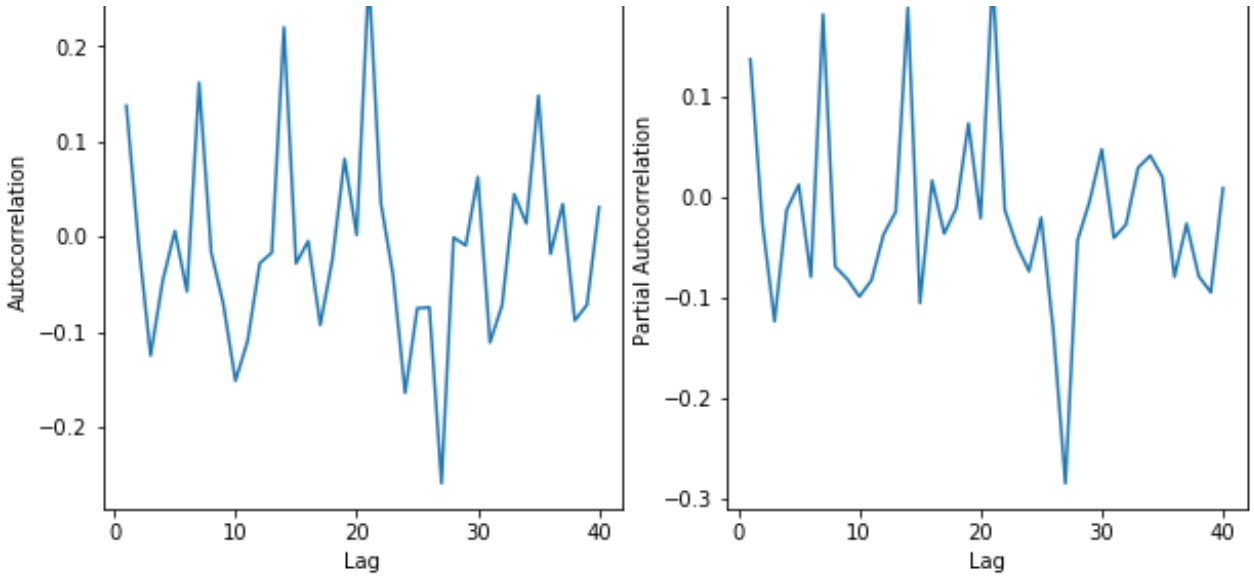
es

	Page	total
92205	Wikipedia:Portada_es.wikipedia.org_all-access_...	751492304.0
95855	Wikipedia:Portada_es.wikipedia.org_mobile-web_...	565077372.0
90810	Especial:Buscar_es.wikipedia.org_all-access_al...	194491245.0
71199	Wikipedia:Portada_es.wikipedia.org_desktop_all...	165439354.0
69939	Especial:Buscar_es.wikipedia.org_desktop_all-a...	160431271.0
94389	Especial:Buscar_es.wikipedia.org_mobile-web_al...	34059966.0
90813	Especial:Entrar_es.wikipedia.org_all-access_al...	33983359.0
143440	Wikipedia:Portada_es.wikipedia.org_all-access_...	31615409.0
93094	Lali_Espósito_es.wikipedia.org_all-access_all-...	26602688.0
69942	Especial:Entrar_es.wikipedia.org_desktop_all-a...	25747141.0

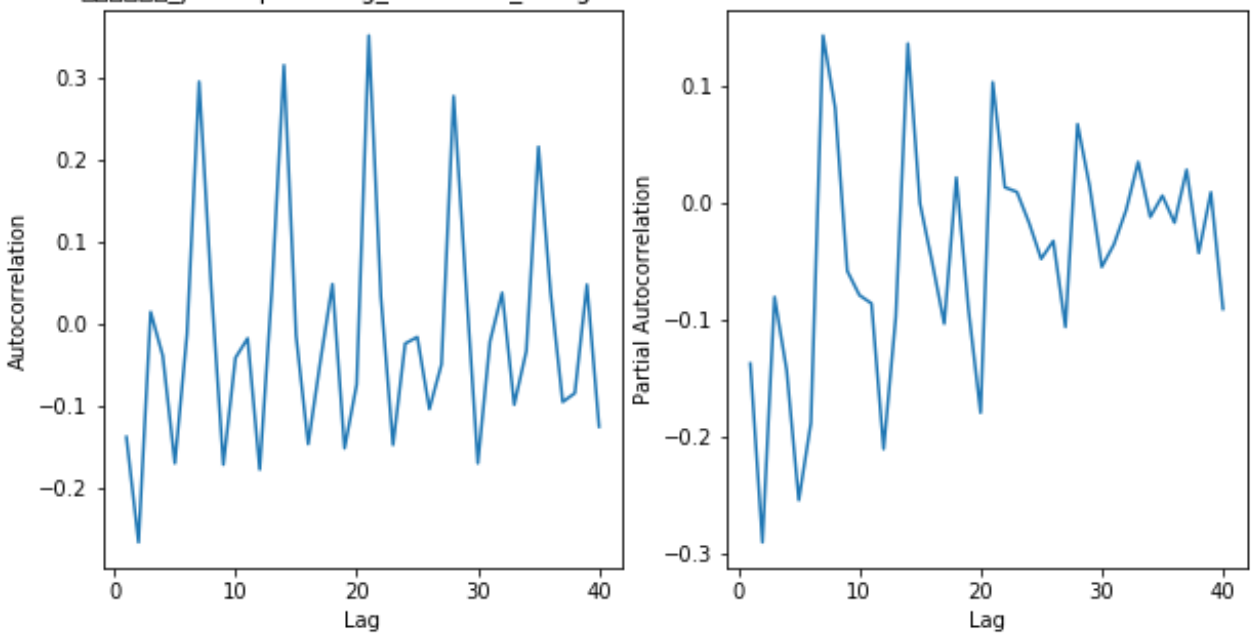
In [8]:

```
from statsmodels.tsa.stattools import pacf
from statsmodels.tsa.stattools import acf

for key in top_pages:
    fig = plt.figure(1,figsize=[10,5])
    ax1 = fig.add_subplot(121)
    ax2 = fig.add_subplot(122)
    cols = train.columns[1:-1]
    data = np.array(train.loc[top_pages[key],cols])
    data_diff = [data[i] - data[i-1] for i in range(1,len(data))]
    autocorr = acf(data_diff)
    pac = pacf(data_diff)

    x = [x for x in range(len(pac))]
    ax1.plot(x[1:],autocorr[1:])

    ax2.plot(x[1:],pac[1:])
    ax1.set_xlabel('Lag')
    ax1.set_ylabel('Autocorrelation')
    ax1.set_title(train.loc[top_pages[key],'Page'])

    ax2.set_xlabel('Lag')
    ax2.set_ylabel('Partial Autocorrelation')
    plt.show()
```

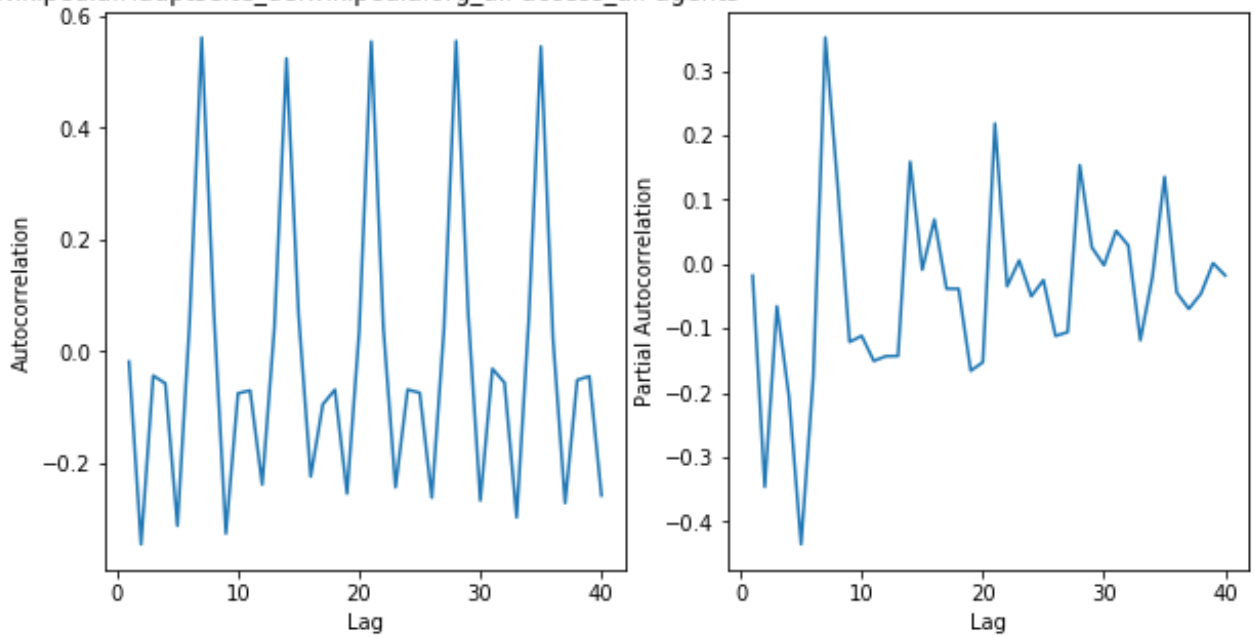




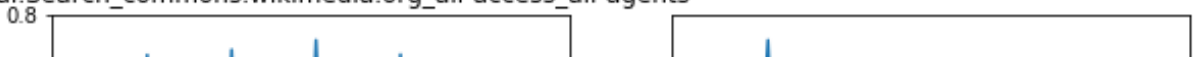
□□□□□□_ja.wikipedia.org_all-access_all-agents

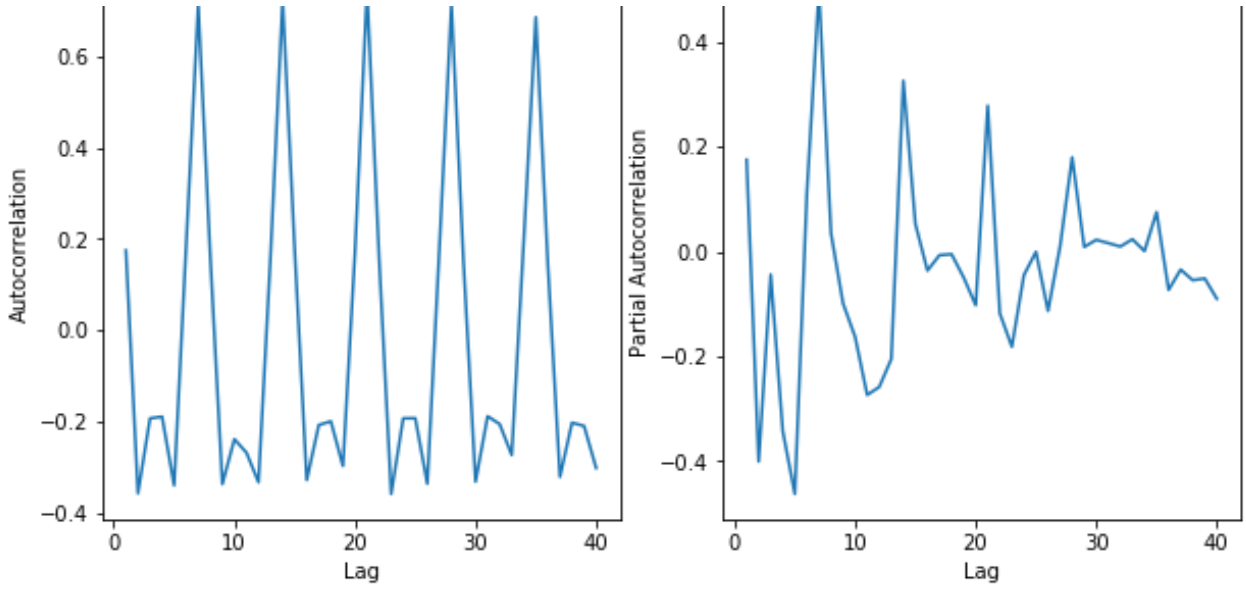


Wikipedia:Hauptseite_de.wikipedia.org_all-access_all-agents

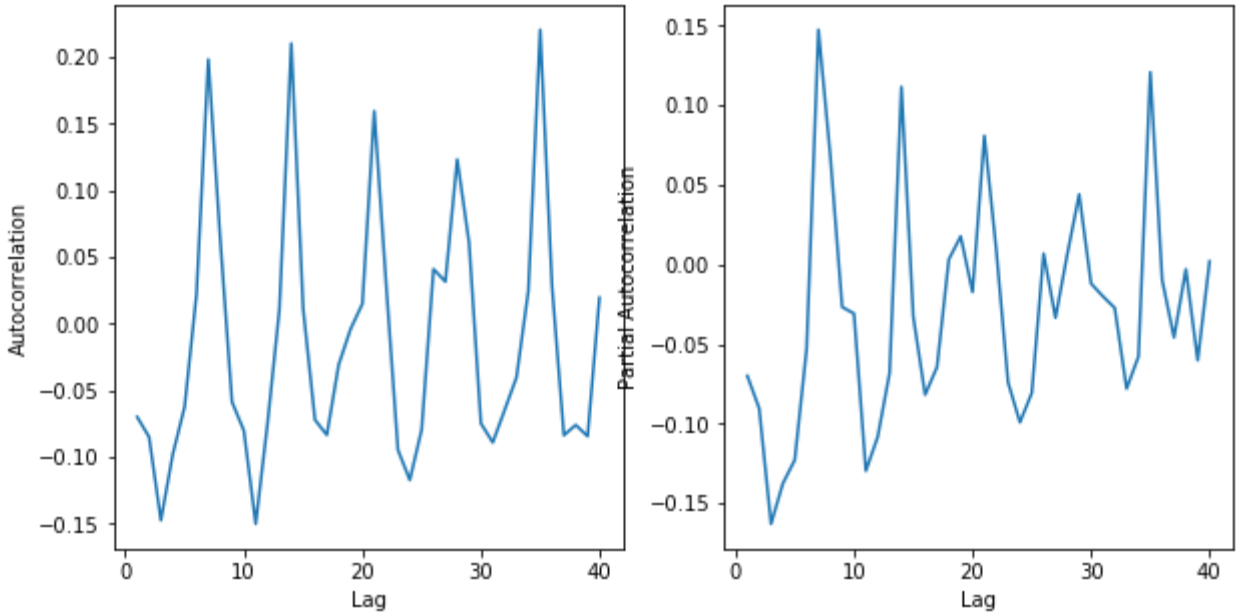


Special:Search_commons.wikimedia.org_all-access_all-agents

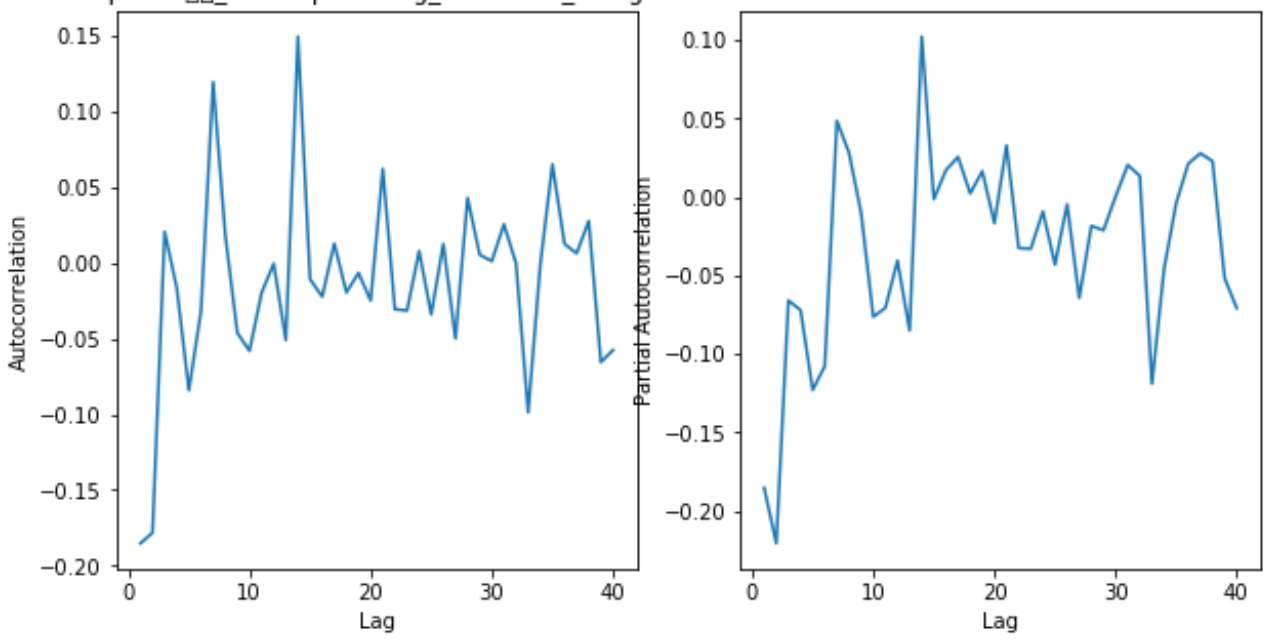




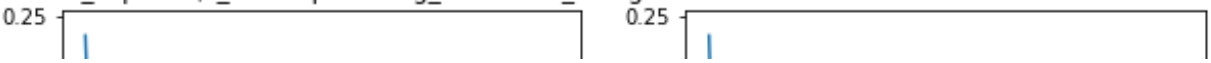
Wikipédia:Accueil_principal_fr.wikipedia.org_all-access_all-agents

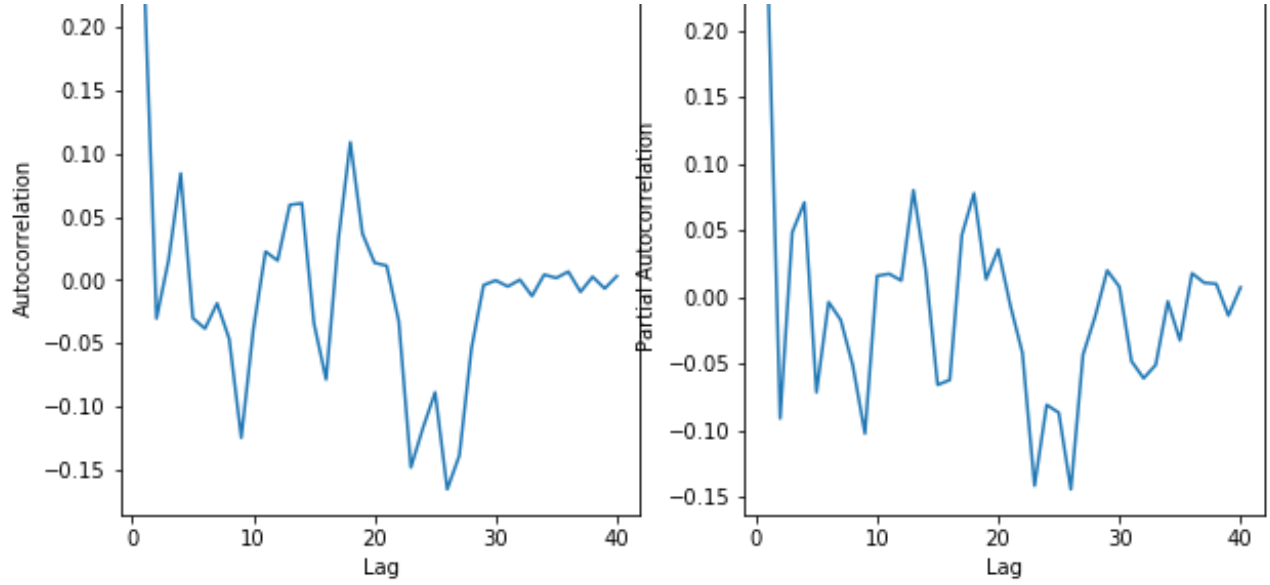


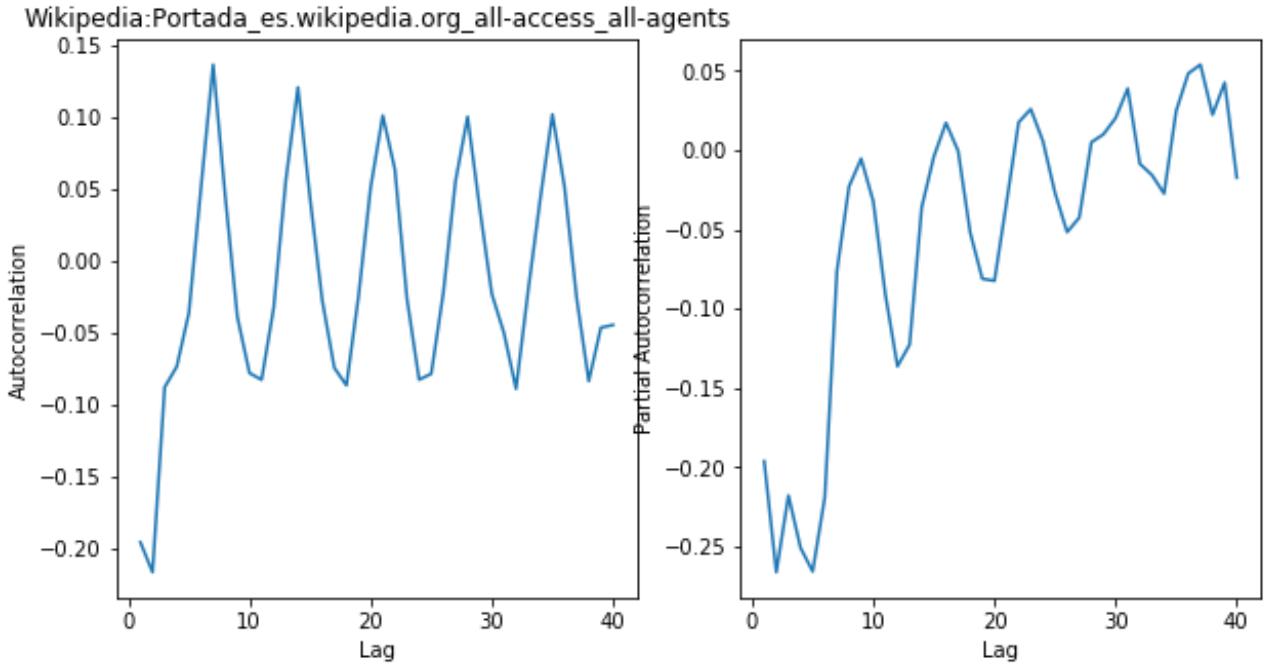
Wikipedia:□□_zh.wikipedia.org_all-access_all-agents



Заглавная_страница_ru.wikipedia.org_all-access_all-agents



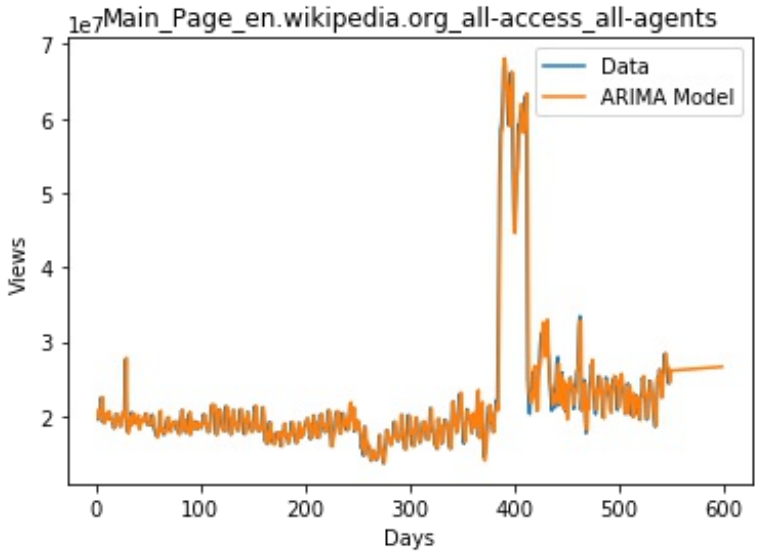


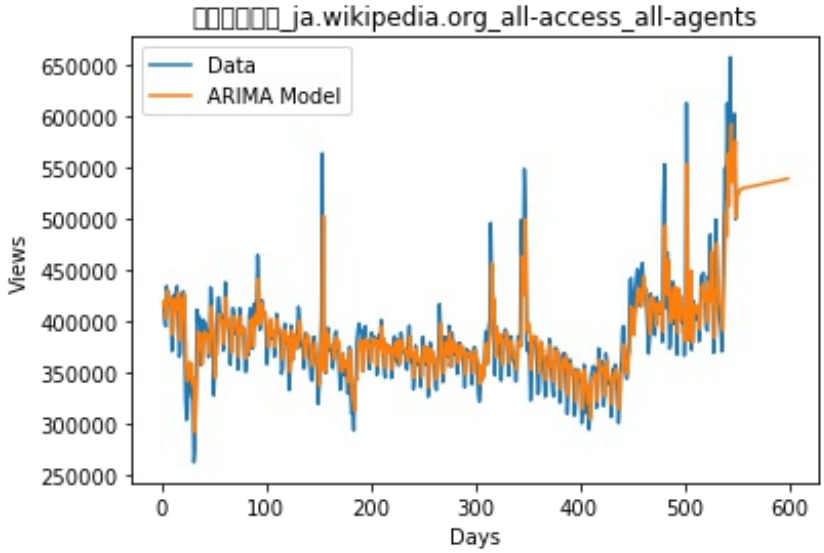


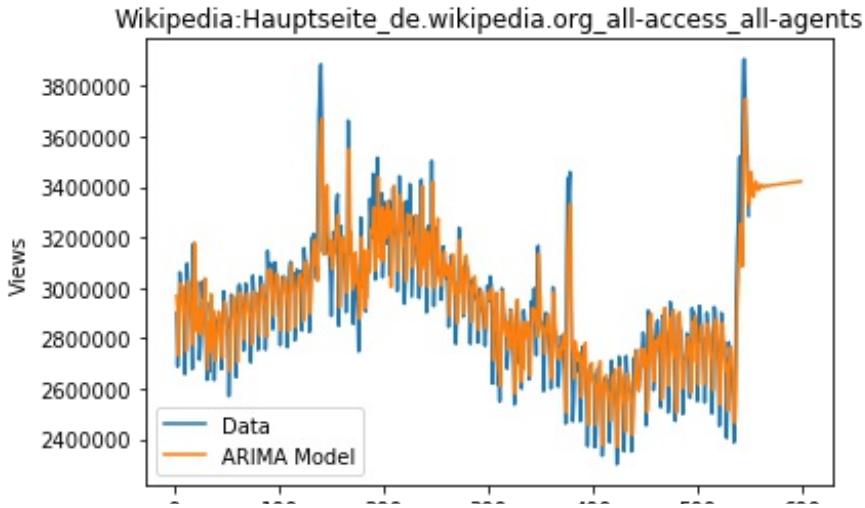
In [9]:

```
from statsmodels.tsa.arima_model import ARIMA
import warnings

cols = train.columns[1:-1]
for key in top_pages:
    data = np.array(train.loc[top_pages[key],cols],'f')
    result = None
    with warnings.catch_warnings():
        warnings.filterwarnings('ignore')
        try:
            arima = ARIMA(data,[2,1,4])
            result = arima.fit(disp=False)
        except:
            try:
                arima = ARIMA(data,[2,1,2])
                result = arima.fit(disp=False)
            except:
                print(train.loc[top_pages[key],'Page'])
                print('\tARIMA failed')
    #print(result.params)
```

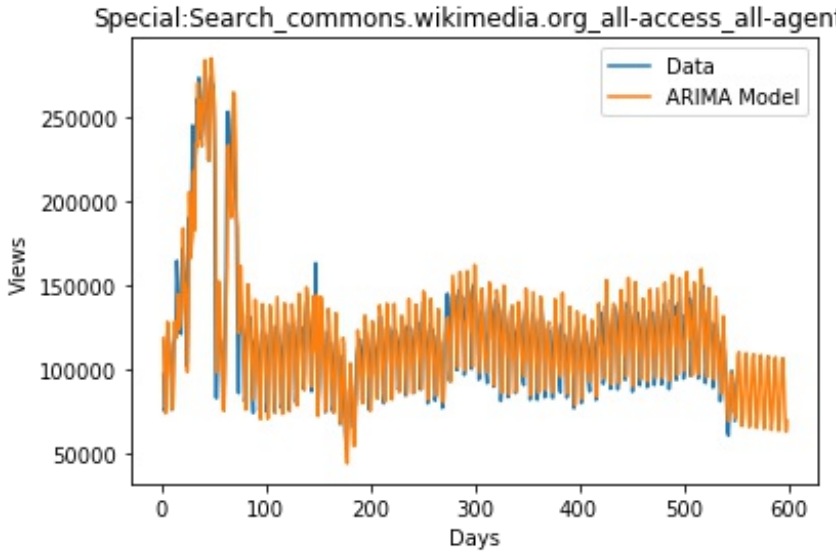
```
    pred = result.predict(2,599,typ='levels')
    x = [i for i in range(600)]
    i=0

    plt.plot(x[2:len(data)],data[2:] ,label='Data')
    plt.plot(x[2:],pred,label='ARIMA Model')
    plt.title(train.loc[top_pages[key],'Page'])
    plt.xlabel('Days')
    plt.ylabel('Views')
    plt.legend()
    plt.show()
```

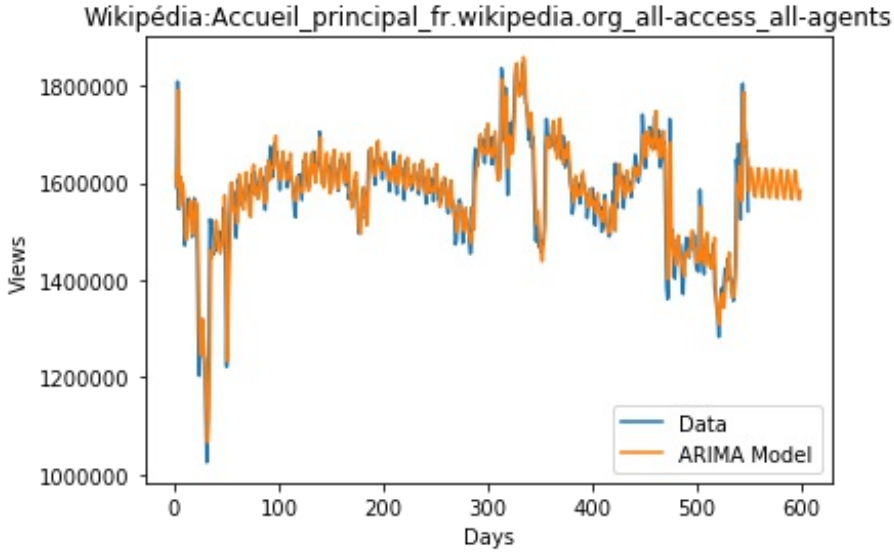


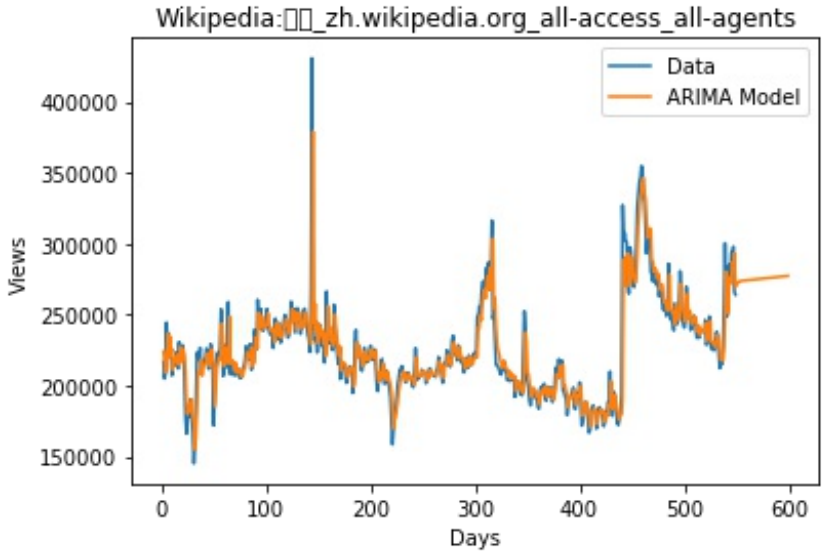




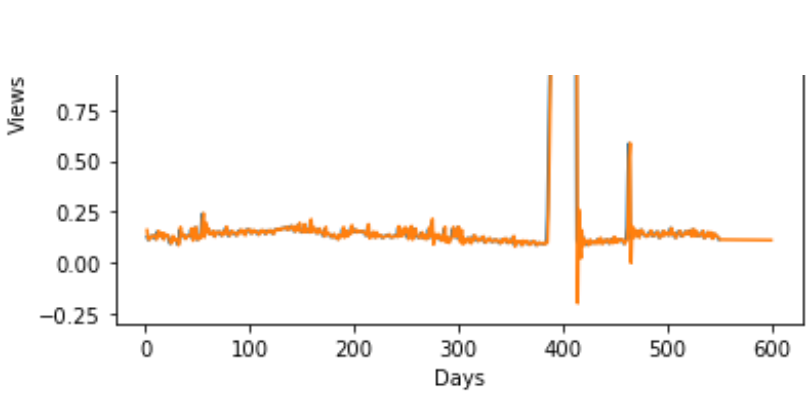

0
100
200
300
400
500
600
Days

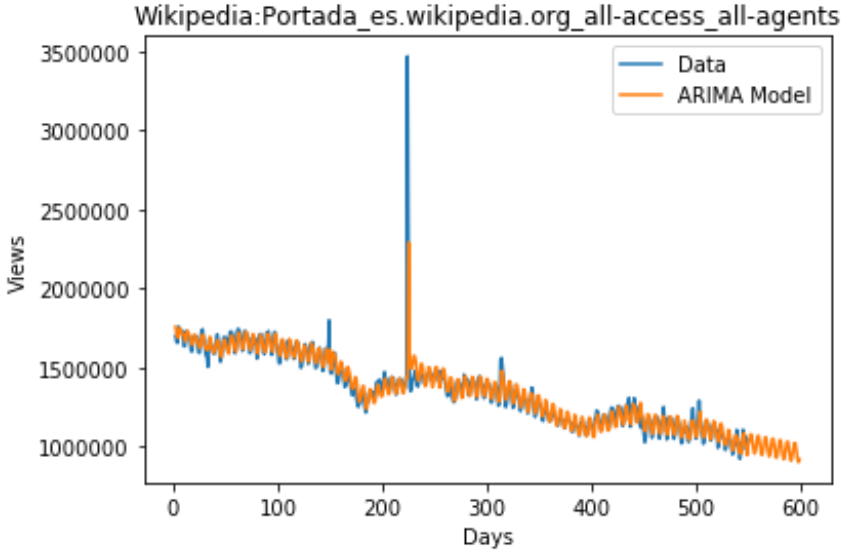

Special:Search_commons.wikimedia.org_all-access_all-agents
Data
ARIMA Model
250000
200000
150000
100000
50000
Views
0
100
200
300
400
500
600
Days


Wikipédia:Accueil_principal_fr.wikipedia.org_all-access_all-agents
1800000
1600000
1400000
1200000
1000000
Views
Data
ARIMA Model
0
100
200
300
400
500
600
Days


Wikipedia:□□_zh.wikipedia.org_all-access_all-agents
400000
350000
300000
250000
200000
150000
Views
Data
ARIMA Model
0
100
200
300
400
500
600
Days


Заглавная_страница_ru.wikipedia.org_all-access_all-agents
1e7
1.75
1.50
1.25
1.00
Data
ARIMA Model


Views
0.75
0.50
0.25
0.00
−0.25
0
100
200
300
400
500
600
Days


Wikipedia:Portada_es.wikipedia.org_all-access_all-agents
Data
ARIMA Model
Views
3500000
3000000
2500000
2000000
1500000
1000000
0
100
200
300
400
500
600
Days