```
import pylab import pandas as pd import numpy as np  df =
pd.read_csv("path/events_train_holdout.tsv",   na_values=['-'],
                    sep = '\t'
                    )
df = df.drop_duplicates() def isNaN(num):
          return num != num

is_not_null_event_id = ~isNaN(df['event_id']) df
= df[is_not_null_event_id]

is_not_nan_start_tstamp = ~isNaN(df['start_tstamp']) temp_start_tstamp_df =
df[is_not_nan_start_tstamp]

temp_start_tstamp_df['year'] = temp_start_tstamp_df['start_tstamp'].astype(str).str[0:4]

is_2008 = temp_start_tstamp_df['year'] == '2008' is_2009 =
temp_start_tstamp_df['year'] == '2009' is_2010 =
temp_start_tstamp_df['year'] == '2010' is_2011 =
temp_start_tstamp_df['year'] == '2011'

#is_year = is_2008 or is_2009 or is_2010 or is_2011

temp_start_tstamp_df_2008 = temp_start_tstamp_df[is_2008] temp_start_tstamp_df_2009
= temp_start_tstamp_df[is_2009] temp_start_tstamp_df_2010 =
temp_start_tstamp_df[is_2010] temp_start_tstamp_df_2011 = temp_start_tstamp_df[is_2011]

temp_start_tstamp_df =
temp_start_tstamp_df_2008.append(temp_start_tstamp_df_2009.append(temp_start_tstamp_df
_2010.append(temp_start_tstamp_df_2011)))

temp_start_tstamp_df['year'] = temp_start_tstamp_df['year'].astype(int)
is_not_nan_confirmed_tstamp = ~isNaN(df['confirmed_tstamp']) temp_confirmed_tstamp_df
= df[is_not_nan_confirmed_tstamp] temp_confirmed_tstamp_df['year'] =
temp_confirmed_tstamp_df['confirmed_tstamp'].astype(str).str[0:4]
is_2003 = temp_confirmed_tstamp_df['year'] == '2003' is_2004 =
temp_confirmed_tstamp_df['year'] == '2004' is_2005 = temp_confirmed_tstamp_df['year']
== '2005' is_2006 = temp_confirmed_tstamp_df['year'] == '2006' is_2007 =
temp_confirmed_tstamp_df['year'] == '2007' is_2008 = temp_confirmed_tstamp_df['year']
== '2008' is_2009 = temp_confirmed_tstamp_df['year'] == '2009' is_2010 =
temp_confirmed_tstamp_df['year'] == '2010' is_2011 = temp_confirmed_tstamp_df['year']
== '2011' is_2012 = temp_confirmed_tstamp_df['year'] == '2012'
#is_year = is_2008 or is_2009 or is_2010 or is_2011
```

```
temp_confirmed_tstamp_df_2003 = temp_confirmed_tstamp_df[is_2003]
temp_confirmed_tstamp_df_2004 = temp_confirmed_tstamp_df[is_2004]
temp_confirmed_tstamp_df_2005 = temp_confirmed_tstamp_df[is_2005]
temp_confirmed_tstamp_df_2006 = temp_confirmed_tstamp_df[is_2006]
temp_confirmed_tstamp_df_2007 = temp_confirmed_tstamp_df[is_2007]
temp_confirmed_tstamp_df_2008 = temp_confirmed_tstamp_df[is_2008]
temp_confirmed_tstamp_df_2009 = temp_confirmed_tstamp_df[is_2009]
temp_confirmed_tstamp_df_2010 = temp_confirmed_tstamp_df[is_2010]
temp_confirmed_tstamp_df_2011 = temp_confirmed_tstamp_df[is_2011]
temp_confirmed_tstamp_df_2012 = temp_confirmed_tstamp_df[is_2012]

temp_confirmed_tstamp_df =
temp_confirmed_tstamp_df_2003.append(temp_confirmed_tstamp_df_2004.append(temp_conf
irmed_tstamp_df_2005.append(temp_confirmed_tstamp_df_2006.append(temp_confirmed_tsta
mp_df_2007.append(temp_confirmed_tstamp_df_2008.append(temp_confirmed_tstamp_df_20
09.append(temp_confirmed_tstamp_df_2010.append(temp_confirmed_tstamp_df_2011.append
(temp_confirmed_tstamp_df_2012)))))))))  is_not_null_event_id =
~isNaN(temp_confirmed_tstamp_df['event_id'])
temp_confirmed_tstamp_df = temp_confirmed_tstamp_df[is_not_null_event_id]

temp_confirmed_tstamp_df['year'] = temp_confirmed_tstamp_df['year'].astype(int)

is_not_nan_created_tstamp = ~isNaN(df['created_tstamp']) temp_created_tstamp_df =
df[is_not_nan_created_tstamp]

temp_created_tstamp_df['year'] = temp_created_tstamp_df['created_tstamp'].astype(str).str[0:4]

is_2003 = temp_created_tstamp_df['year'] == '2003' is_2004 =
temp_created_tstamp_df['year'] == '2004' is_2005 =
temp_created_tstamp_df['year'] == '2005' is_2006 =
temp_created_tstamp_df['year'] == '2006' is_2007 =
temp_created_tstamp_df['year'] == '2007' is_2008 =
temp_created_tstamp_df['year'] == '2008' is_2009 =
temp_created_tstamp_df['year'] == '2009' is_2010 =
temp_created_tstamp_df['year'] == '2010' is_2011 =
temp_created_tstamp_df['year'] == '2011' is_2012 =
temp_created_tstamp_df['year'] == '2012' #is_year = is_2008
or is_2009 or is_2010 or is_2011

temp_created_tstamp_df_2003 = temp_created_tstamp_df[is_2003]
temp_created_tstamp_df_2004 = temp_created_tstamp_df[is_2004]
temp_created_tstamp_df_2005 = temp_created_tstamp_df[is_2005]
temp_created_tstamp_df_2006 = temp_created_tstamp_df[is_2006]
temp_created_tstamp_df_2007 = temp_created_tstamp_df[is_2007]
temp_created_tstamp_df_2008 = temp_created_tstamp_df[is_2008]
```

```
temp_created_tstamp_df_2009 = temp_created_tstamp_df[is_2009]
temp_created_tstamp_df_2010 = temp_created_tstamp_df[is_2010]
temp_created_tstamp_df_2011 = temp_created_tstamp_df[is_2011]
temp_created_tstamp_df_2012 = temp_created_tstamp_df[is_2012]

temp_created_tstamp_df =
 temp_created_tstamp_df_2003.append(temp_created_tstamp_df_2004.append(temp_created_
                                                                    t
stamp_df_2005.append(temp_created_tstamp_df_2006.append(temp_created_tstamp_df_2007
.append(temp_created_tstamp_df_2008.append(temp_created_tstamp_df_2009.append(temp_
created_tstamp_df_2010.append(temp_created_tstamp_df_2011.append(temp_created_tstamp
_df_2012)))))))) is_not_null_event_id =
~isNaN(temp_created_tstamp_df['event_id']) temp_created_tstamp_df =
temp_created_tstamp_df[is_not_null_event_id] temp_created_tstamp_df['year']
=
temp_created_tstamp_df['year'].astype(int)

is_not_nan_closed_tstamp = ~isNaN(df['closed_tstamp']) temp_closed_tstamp_df =
df[is_not_nan_closed_tstamp]

temp_closed_tstamp_df['year'] = temp_closed_tstamp_df['closed_tstamp'].astype(str).str[0:4]

is_2003 = (temp_closed_tstamp_df['year'] == '2003') is_2004
= (temp_closed_tstamp_df['year'] == '2004') is_2005 =
(temp_closed_tstamp_df['year'] == '2005') is_2006 =
(temp_closed_tstamp_df['year'] == '2006') is_2007 =
(temp_closed_tstamp_df['year'] == '2007') is_2008 =
(temp_closed_tstamp_df['year'] == '2008') is_2009 =
(temp_closed_tstamp_df['year'] == '2009') is_2010 = (temp_closed_tstamp_df['year']
== '2010') is_2011 =
(temp_closed_tstamp_df['year'] == '2011') is_2012 =
(temp_closed_tstamp_df['year'] == '2012') is_2013 =
(temp_closed_tstamp_df['year'] == '2013')

temp_closed_tstamp_df_2003 = temp_closed_tstamp_df[is_2003]
temp_closed_tstamp_df_2004 = temp_closed_tstamp_df[is_2004]
temp_closed_tstamp_df_2005 = temp_closed_tstamp_df[is_2005]
temp_closed_tstamp_df_2006 = temp_closed_tstamp_df[is_2006]
temp_closed_tstamp_df_2007 = temp_closed_tstamp_df[is_2007]
temp_closed_tstamp_df_2008 = temp_closed_tstamp_df[is_2008]
temp_closed_tstamp_df_2009 = temp_closed_tstamp_df[is_2009]
temp_closed_tstamp_df_2010 = temp_closed_tstamp_df[is_2010]
temp_closed_tstamp_df_2011 = temp_closed_tstamp_df[is_2011]
```

```
temp_closed_tstamp_df_2012 = temp_closed_tstamp_df[is_2012]
temp_closed_tstamp_df_2013 = temp_closed_tstamp_df[is_2013]

temp_closed_tstamp_df =
temp_closed_tstamp_df_2003.append(temp_closed_tstamp_df_2004.append(temp_closed_tst
a
mp_df_2005.append(temp_closed_tstamp_df_2006.append(temp_closed_tstamp_df_2007.app
end(temp_closed_tstamp_df_2008.append(temp_closed_tstamp_df_2009.append(temp_closed
_tstamp_df_2010.append(temp_closed_tstamp_df_2011.append(temp_closed_tstamp_df_2012
.append(temp_closed_tstamp_df_2013)))))))))  temp_closed_tstamp_df['year']
= temp_closed_tstamp_df['year'].astype(int)

temp_df =
temp_start_tstamp_df.append(temp_confirmed_tstamp_df.append(temp_created_tstamp_df.ap
pend(temp_closed_tstamp_df)))

df['year'] = np.nan

df_no_duplicates = df.drop_duplicates('event_id')

temp_df_no_duplicates = temp_df.drop_duplicates('event_id')

%matplotlib inline year_grouped = temp_df_no_duplicates.groupby('year')
year_grouped['event_id'].nunique().plot(kind='bar')
```