In [1]

import numpy as np *# linear algebra*

import pandas as pd *# data processing, CSV file I/O (e.g. pd.read\_csv)*

import os

print(os.listdir("../input"))

In[2]

cols\_keep = ['Accident\_Severity', 'Date','Time', 'Latitude','Longitude',

'Local\_Authority\_(District)', 'Local\_Authority\_(Highway)',

'LSOA\_of\_Accident\_Location', 'Number\_of\_Casualties', "1st\_Road\_Number","2nd\_Road\_Number"]

In[3]

df = pd.read\_csv('../input/Accident\_Information.csv',usecols=cols\_keep, *#nrows=12345,*

parse\_dates=[['Date', 'Time']],keep\_date\_col=True) df.shape

In[4]

df["Date\_Time"]=

pd.to\_datetime(df["Date\_Time"],infer\_datetime\_format=True,errors="coerce")

In[5]

*# we see that some cases lack a time of events - creating a bad date format. we'll fix these*

df.loc[df['Date\_Time'].isna(), 'Date\_Time'] = df["Date"]

df.loc[df["Date\_Time"].isna()]

In[6]

df.drop(["Date","Time"],axis=1,inplace=True)

df.set\_index("Date\_Time",inplace=True)

df.index = pd.to\_datetime(df.index)

In[7]

df["serious\_accident"] = df.Accident\_Severity != "Slight

In[8]

df.nunique()

In[9]

df.columns

In[10]

df.describe()

In[11]

df.index.dtype

In[12]

df.head()

In[13]

*# Identifying the worst districts to travel.*

*### https://stackoverflow.com/questions/19384532/how-to-count-number-of-rows-per-group-and-other-statistics-in-pandas-group-by*

*###* [*https://stackoverflow.com/questions/32012012/pandas-resample-timeseries-with-groupby/39186403#39186403*](https://stackoverflow.com/questions/32012012/pandas-resample-timeseries-with-groupby/39186403#39186403)

lsoa\_wise = df.groupby( 'LSOA\_of\_Accident\_Location').resample("M").agg({"Number\_of\_Casualties":"sum","serious\_accident":"sum",

"Accident\_Severity":"count",

*# "Latitude":scipy.stats.mode,"Longitude":scipy.stats.mode*

*# "Latitude":"mean","Longitude":"mean"*

*# we get missing latLong when no accidents occured, and their locations can change unless we use mode!* })

lsoa\_wise.rename(columns={"Accident\_Severity":"Accident\_counts"},inplace=True)

lsoa\_wise["percent\_seriousAccidents"] = 100\*lsoa\_wise["serious\_accident"]/lsoa\_wise["Accident\_counts"].round(2)

lsoa\_wise.loc[lsoa\_wise['percent\_seriousAccidents'].isna(), 'percent\_seriousAccidents'] = 0

print(lsoa\_wise.shape)

lsoa\_wise.head()

In[14]

lsoa\_wise.describe()

In[15]

lsoa\_wise.to\_csv("uk\_accidents\_lsoa\_monthly.csv.gz",compression="gzip")