



فراخوانی چهار فایل قبلی CSV با استفاده از Pandas

چاپ تعداد ردیف و ستون فایل ها

بررسی نوع داده / تبدیل ستون های تاریخ و عددی به Float , Datetime

پیدا کردن مقادیر گمشده / جایگزین یا حذف

ایجاد ستون جدیدی به نام Month از روی time

ایجاد ستون Category برای طبقه بندی شدت زلزله

گروه بندی داده ها براساس Month , Category محاسبه میانگین و تعداد زلزله ها

ایجاد ستون Region یا Area\_name یا Place با استخراج از ستون که نشان دهنده منطقه وقوع زلزله می باشد و انجام آن برای تمامی دیتا

گروه بندی داده ها براساس منطقه تحلیل آماری با استفاده از متد Groupby

تعداد زلزله در هر منطقه با Size یا Count

میانگین بزرگی و عمق زلزله ها در هر منطقه حداکثر بزرگی یا عمق ثبت شده در هر منطقه

رسم نمودار میله ای از تعداد زلزله ها در هر منطقه با استفاده از خروجی Groupby

محاسبه فاصله کانونی هر کدام از زلزله ها تا توکیو پایتخت ژاپن

ایجاد ستون برای ذخیره سازی دیتا

انجام محاسبات آماری و ریاضی روی آرایه ها

ایجاد جداول در دیتابیس ایجاد جدولی به نام Earthquakes

ایجاد ستون های ذیل در جدول قبلی ایجاد شده:

id

Time

latitude, longitude

Depth

Magnitude

Region

Source منبع داده آن: USGS , GEOFON

با استفاده از Pandas.to-sql دیتا های تمیز شده را از DataFrame به جدول SQL انتقال می دهیم.

استفاده از Chunksize و مشخص نمودن 'if\_exists='append

نوشتن کوئری ها در اسکیل

- محاسبه تعداد کل زلزله ها به تفکیک ماه و منطقه
- محاسبه میانگین بزرگی در هر منطقه و منبع
- استخراج ۱۰ زلزله شدید اخیر
- محاسبه بیشترین و کمترین عمق در هر منطقه
- حذف داده های خارج از محدوده یا مشکوک
- به روزرسانی داده های ناقص
- تعریف شاخص ایندکس برای ستون های پر جست و جو مانند Region , Time , Magnitude برای بهبود سرعت کوئری ها

مصور سازی داده ها:

- استفاده از کتابخانه Matplotlib
- رسم نمودارهای هیستوگرام خطی پراکنندگی جمعیه ای هیت مپ

تست نویسی و اعتبارسنجی:

نوشتن تست واحد برای اطمینان از صحت مراحل انجام شده:

- بررسی صحت دریافت داده
- اطمینان از حذف مقادیر ناقص
- صحت نوع داده ها
- محاسبات آماری
- صحت درج داده ها در پایگاه داده

نتیجه گیری و تحلیل نهایی:

تفاوت های شاخص بین این زلزله ها

عمق زلزله ها / شدت آن ها / رتبه بندی خطرناک بودن این زلزله ها

- بررسی پوشش تمام منابع رخدادهای کوچک، تفاوت آن ها با یکدیگر
- بررسی اینکه زلزله های شدید در چه عمقی هستند
- رتبه بندی خطرناکترین زلزله ها
- مقایسه تعداد زلزله های بزرگ در هر منبع
- نتیجه گیری علمی درباره رفتار زلزله های ژاپن در ماه اخیر.
- پیشنهاد برای ترکیب منابع برای بهینه سازی کیفیت داده ها.

پایان