https://www.kaggle.com/code/dsyt17/notebook56e9822521/edit

1. Описание бизнес-процесса, выявление проблем и потребностей бизнес-анализа.

Изучение землетрясений позволяет подготовиться к ним лучше в тех местах, где это необходимо.

2. Формирование исходных структурированных данных, разработка структуры хранилища данных

для анализа, консолидация данных.

**Data columns (total 10 columns):**

# Column Non-Null Count Dtype

0 Date 14344 non-null object

1 Time (utc) 14344 non-null object

2 Region 14344 non-null object

3 Magnitude 14344 non-null float64

4 Depth (km) 14344 non-null int64

5 Latitude 14344 non-null float64

6 Longitude 14344 non-null float64

7 Mode 14344 non-null object

8 Map 14344 non-null object

9 Datetime 14344 non-null datetime64[ns]

dtypes: datetime64[ns](1), float64(3), int64(1), object(5)

Выкидываем все строки с пустыми значениями:

df.replace('', np.nan, inplace=True)

df.dropna(subset=['Date', 'Time (utc)', 'Magnitude', 'Depth (km)', 'Latitude', 'Longitude'], inplace=True)

Преобразование столбцов даты и времени в datetime, преобразование столбцов величины и глубины в числовые типы, преобразование широты и долготы в плавающие значения после удаления указателей направления

df['Datetime'] = pd.to\_datetime(df['Date'] + ' ' + df['Time (utc)'], format='**%d**/%m/%Y %H:%M:%S')

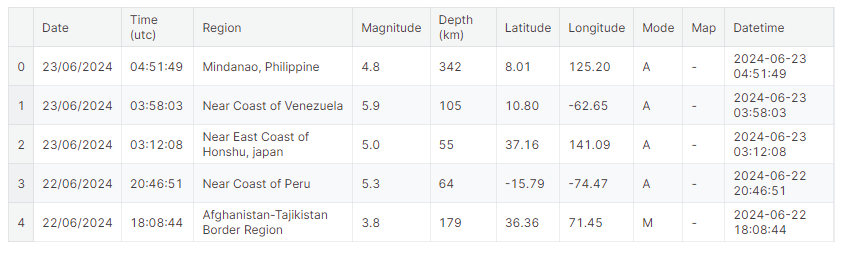
df['Magnitude'] = pd.to\_numeric(df['Magnitude'], errors='coerce')

df['Depth (km)'] = pd.to\_numeric(df['Depth (km)'], errors='coerce')

df['Latitude'] = df['Latitude'].str.extract(r'(\d+\.\d+)')[0].astype(float) \* df['Latitude'].str.extract(r'([NS])')[0].map({'N': 1, 'S': -1})

df['Longitude'] = df['Longitude'].str.extract(r'(\d+\.\d+)')[0].astype(float) \* df['Longitude'].str.extract(r'([EW])')[0].map({'E': 1, 'W': -1})

Данные принимают следующий вид:



3. Разработка сценариев анализа и формирование аналитической OLAP-отчетности.

