# Analysis of Mineral Ores

Yahya

EDA project

21/11/2021

# Outline:

- Reading dataset into data frames
- Querying database using SQL statements within python
- Plotting graphs

# Data Structures

```
dataset = pd.read_csv("mineral_dataset.csv", sep=',', usecols=[0,1,2,4,8,20])
dataset= dataset.dropna()
dataset.head()
```

| | site_name | latitude | longitude | country | commod1 | hrock_type |
|---|---|---|---|---|---|---|
| 0 | Lookout Prospect | 55.05612 | -132.14344 | United States | Copper | Schist |
| 1 | Lucky Find Prospect | 55.52751 | -132.68514 | United States | Copper | Diabase |
| 2 | Mccullough Prospect | 55.97751 | -132.99906 | United States | Copper | Siltstone |
| 3 | Lucky Jim Claim | 55.52195 | -132.68653 | United States | Gold | Granite |
| 4 | Matilda Occurrence | 55.14556 | -132.05233 | United States | Gold | Mica Schist |

```
dataset.shape
```

```
(64307, 6)
```

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 64307 entries, 0 to 304341
Data columns (total 6 columns):
site_name     64307 non-null object
latitude      64307 non-null float64
longitude     64307 non-null float64
country       64307 non-null object
commod1       64307 non-null object
hrock_type    64307 non-null object
dtypes: float64(2), object(4)
memory usage: 3.4+ MB
```

# Data Structures

```
from sqlalchemy import create_engine
engine = create_engine('sqlite:///top10_rocktype_v3.db', echo=False)

rocktype = {key: int(value) for key, value in
         engine.execute("SELECT hrock_type, Size  FROM hrock_type \
         ORDER BY Size DESC LIMIT 7" ).fetchall()}
rocktype
```

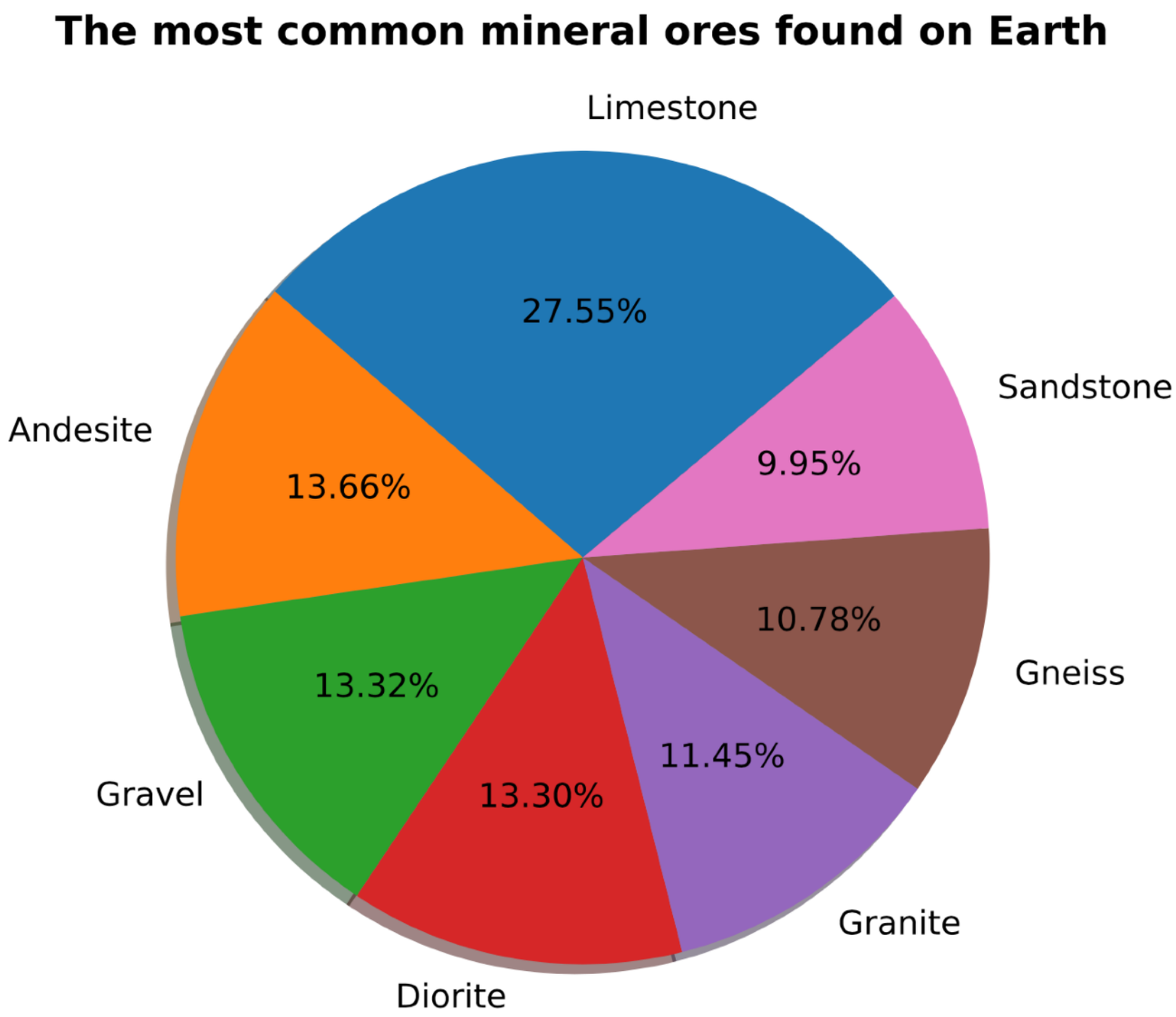
```
{'Limestone': 6423,
 'Andesite': 3184,
 'Gravel': 3105,
 'Diorite': 3101,
 'Granite': 2669,
 'Gneiss': 2514,
 'Sandstone': 2321}
```

```
rck = ['Limestone','Andesite','Gravel','Diorite']
countries = ["United States","Mexico", "Peru","Chile" ,"Argentina"]
newDF = pd.concat([rocktype_group.get_group(i) for i in rck])
```

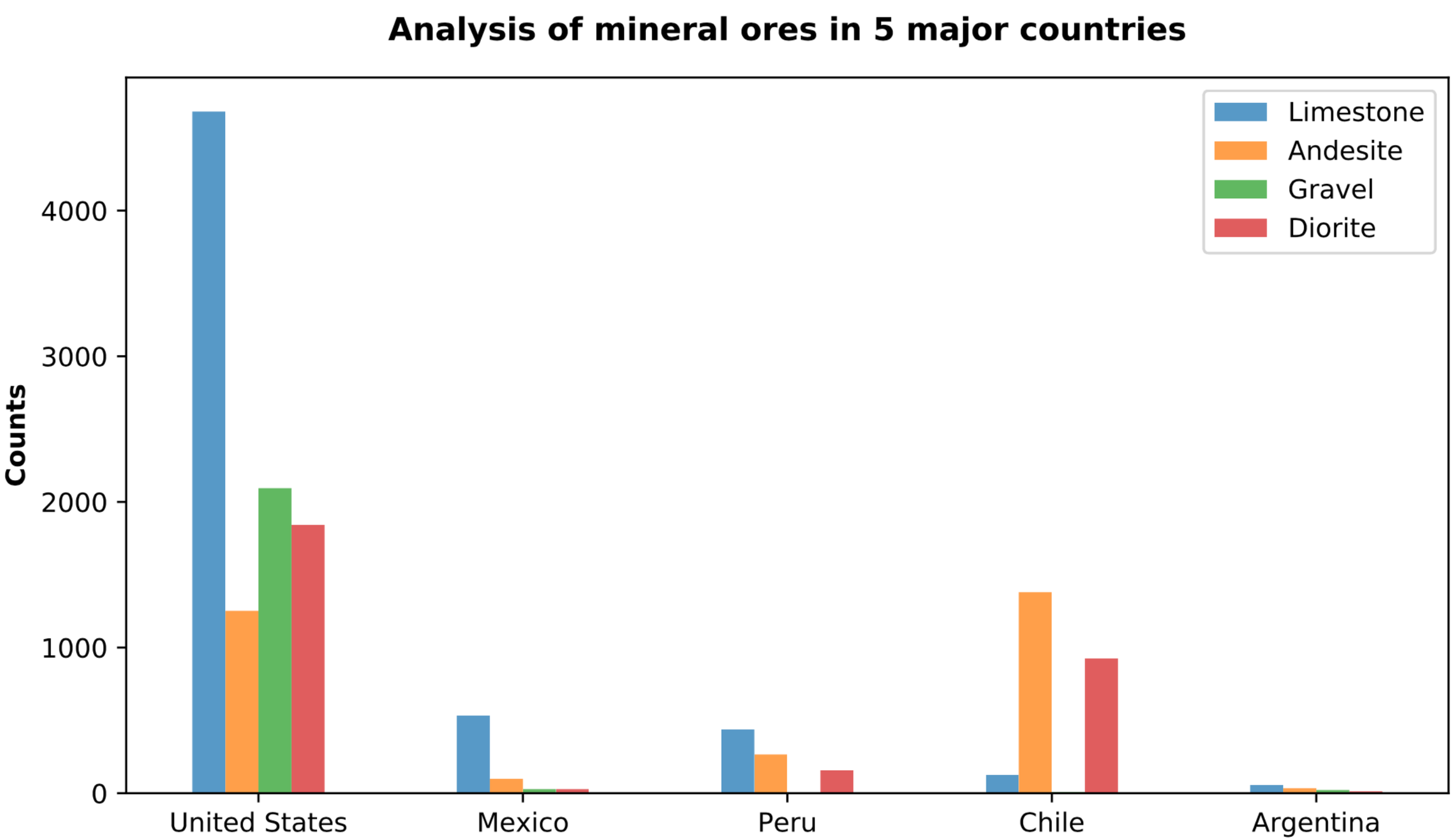
```
rock_cnt = {i:  {j: newDF[(newDF['country'] == j) & (newDF['hrock_type']== i) ].count()[-1] \
     for j in countries} for i in rck}
[*rock_cnt.values()][0]
```

```
{'United States': 4680,
 'Mexico': 535,
 'Peru': 439,
 'Chile': 127,
 'Argentina': 58}
```

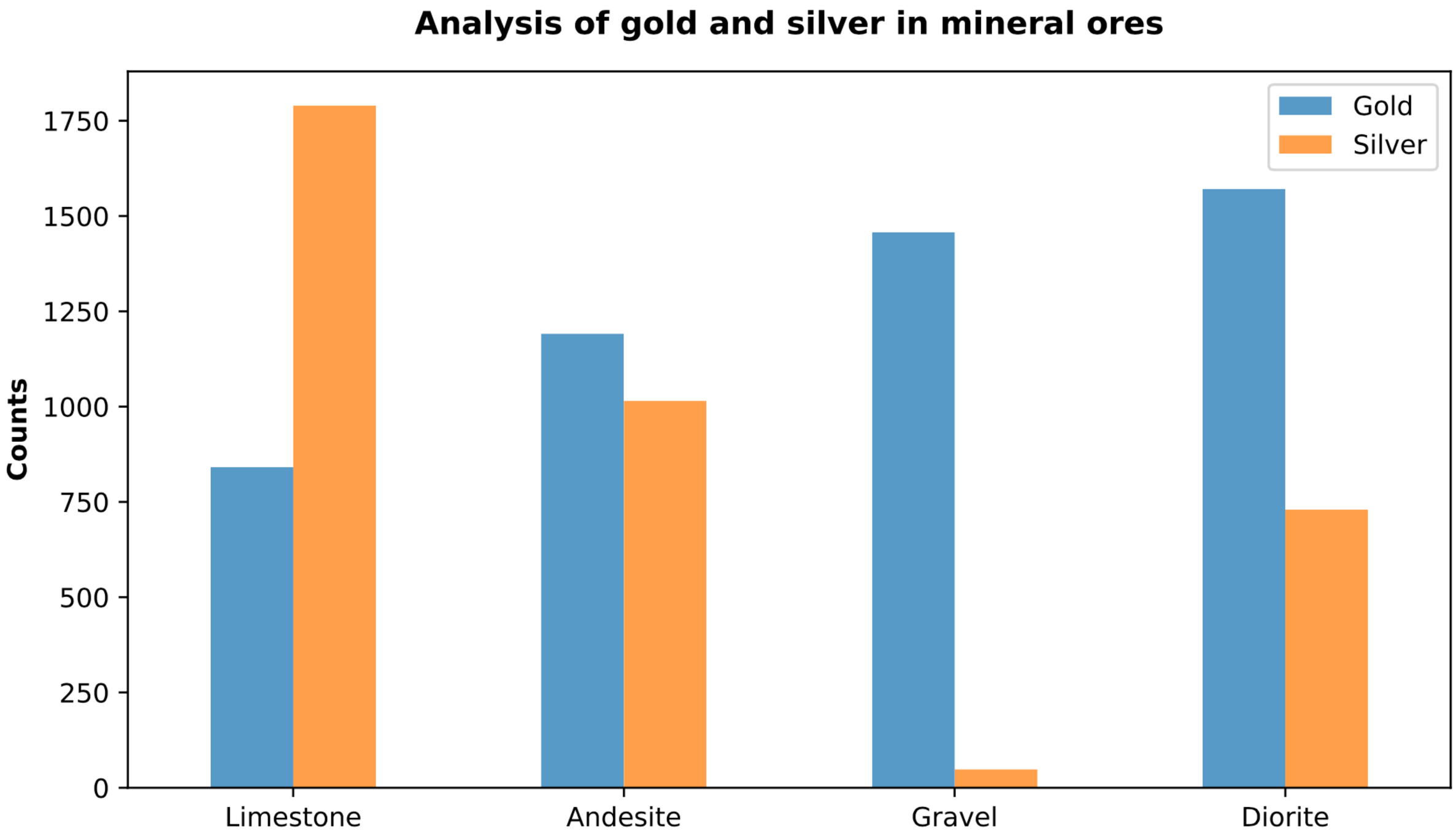
# Data Visualization



# Data Visualization



# Data Visualization



# Thanks!

## Any Questions?