

Exploratory Data Analysis ## Data Schema

- Data comes in from many sources
- Once you connect data to S3, you can bring it into Sage Maker

Data source ~ breast cancer wisconsin data set

Data Schema: Merge / Joins

Types: categorical, ordinal, numerical, date, vector, text, image, unstructured, etc.

Use pandas DataFrame merge/join to join two data sets

```
df = pd.DataFrame(...)
```

```
df2 = pd.DataFrame(...)
```

```
df.merge(df2, on = "Job", how = 'inner')
```