```
import pandas as pd
import numpy as np

from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
df =
pd.read_csv("/content/drive/MyDrive/Foldah/Meteorite_Landings.csv")
```

Using the meteorite data from the Meteorite_Landings.csv file, update the year column to only contain the year, convert it to a numeric data type, and create a new column indicating whether the meteorite was observed falling before 1970. Set the index to the id column and extract all the rows with IDs between 10,036 and 10,040 (inclusive) with loc[].

Hint 1: Use year.str.slice() to grab a substring.

Hint 2: Make sure to sort the index before using loc[] to select the range.

Bonus: There's a data entry error in the year column. Can you find it? (Don't spend too much time on this.)

```
df['year'] = df ['year'].str.slice(6,11)
df.head(3)
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 45716,\n  \"fields\":
[\n    {\n      \"column\": \"name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 45716,\n
\"samples\": [\n          \"Grove Mountains 024259\",\n
\"LaPaz Icefield 02382\",\n          \"Yamato 86722\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"id\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 16860,\n        \"min\": 1,\n
\"max\": 57458,\n        \"num_unique_values\": 45716,\n
```

\"samples\": [\n   50216,\n   12649,\n   30228\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"nametype\",\n
\"properties\": {\n   \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n   \"samples\": [\n
\"Relict\",\n   \"Valid\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"recclass\",\n   \"properties\":
{\n   \"dtype\": \"category\",\n   \"num_unique_values\":
466,\n   \"samples\": [\n   \"H5-6\",\n
\"CO3.3\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"mass (g)\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 574988.87641047,\n   \"min\": 0.0,\
n   \"max\": 60000000.0,\n   \"num_unique_values\": 12576,\n
\"samples\": [\n   1521.1,\n   56.16\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"fall\",\n   \"properties\": {\n
\"dtype\": \"category\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   \"Found\",\n   \"Fell\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"year\",\n   \"properties\": {\n
\"dtype\": \"category\",\n   \"num_unique_values\": 265,\n
\"samples\": [\n   \"1857 \",\n   \"1861 \"\n   ],\
n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"reclat\",\n   \"properties\":
{\n   \"dtype\": \"number\",\n   \"std\":
46.37851135669297,\n   \"min\": -87.36667,\n   \"max\":
81.16667,\n   \"num_unique_values\": 12738,\n   \"samples\":
[\n   21.06917,\n   20.53877\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"reclong\",\n   \"properties\":
{\n   \"dtype\": \"number\",\n   \"std\":
80.64729807906366,\n   \"min\": -165.43333,\n   \"max\":
354.47333,\n   \"num_unique_values\": 14640,\n
\"samples\": [\n   54.70452,\n   161.37957\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"GeoLocation\",\n
\"properties\": {\n   \"dtype\": \"category\",\n
\"num_unique_values\": 17100,\n   \"samples\": [\n
\"(18.58833, 54.01833)\",\n   \"(-72.77778, 75.32639)\"\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   }\n]\n}","type":"dataframe","variable_name":"df"}

```
meteorite.dtypes
```

```
name            object
id               int64
nametype        object
recclass        object
```

```
mass (g)       float64
fall            object
year            object
reclat         float64
reclong        float64
GeoLocation     object
dtype: object
```

```
import pandas as pd

df['year'] = df['year'].str.slice(6, 10)

df['year'] = pd.to_numeric(df['year'], errors='coerce')
df['observed_fell_before_1970'] = (df['fall'] == 'Fell') & (df['year']
< 1970)

df.set_index('id', inplace=True)

df.sort_index(inplace=True)

extracted_rows = df.loc[10036:10040]

print("Extracted Rows (IDs 10036 to 10040):")
print(extracted_rows)

data_error = df[df['year'] > 2024]
print("\nData Entry Error Found:")
print(data_error[['name', 'year']])
```

```
Extracted Rows (IDs 10036 to 10040):
           name nametype         recclass  mass (g)   fall  year
reclat  \
id

10036     Enigma    Valid               H4      94.0  Found   NaN
31.33333
10037       Enon    Valid  Iron, ungrouped     763.0  Found   NaN
39.86667
10038      Enshi    Valid               H5    8000.0   Fell   NaN
30.30000
10039  Ensisheim    Valid              LL6  127000.0   Fell   NaN
47.86667

          reclong            GeoLocation  observed_fell_before_1970
id
10036  -82.31667  (31.33333, -82.31667)                       False
10037  -83.95000     (39.86667, -83.95)                       False
10038  109.50000          (30.3, 109.5)                       False
10039    7.35000       (47.86667, 7.35)                       False

Data Entry Error Found:
```

```
Empty DataFrame
Columns: [name, year]
Index: []
```