cancer_test_results

January 21, 2019

1 Cancer Test Results

```
In [1]: import numpy as np
        import pandas as pd
        # load dataset
        cancer_df = pd.read_csv('cancer_test_data.csv')
        cancer_df.head()
```

```
Out[1]:    patient_id test_result  has_cancer
        0       79452    Negative       False
        1       81667    Positive        True
        2       76297    Negative       False
        3       36593    Negative       False
        4       53717    Negative       False
```

```
In [5]: # number of patients
        no = len(cancer_df)
        no
```

```
Out[5]: 2914
```

```
In [6]: # number of patients with cancer
        no_cancer = len(cancer_df[cancer_df['has_cancer'] == True])
        no_cancer
```

```
Out[6]: 306
```

```
In [7]: # number of patients without cancer
        no_notcancer = len(cancer_df[cancer_df['has_cancer'] == False])
        no_notcancer
```

```
Out[7]: 2608
```

```
In [8]: # proportion of patients with cancer
        prop_cancer = no_cancer/no
        prop_cancer
```

```
Out[8]: 0.10501029512697323
```

```
In [10]:  # proportion of patients without cancer
          prop_notcancer = 1 - prop_cancer
          prop_notcancer

Out[10]:  0.8949897048730268

In [12]:  # proportion of patients with cancer who test positive
          (cancer_df.query('has_cancer')['test_result'] == 'Positive').mean()

Out[12]:  0.90522875816993464

In [13]:  # proportion of patients with cancer who test negative
          (cancer_df.query('has_cancer')['test_result'] == 'Negative').mean()

Out[13]:  0.094771241830065356

In [14]:  # proportion of patients without cancer who test positive
          (cancer_df.query('has_cancer == False')['test_result'] == 'Positive').mean()

Out[14]:  0.2036042944785276

In [15]:  # proportion of patients without cancer who test negative
          (cancer_df.query('has_cancer == False')['test_result'] == 'Negative').mean()

Out[15]:  0.79639570552147243

In [ ]:
```