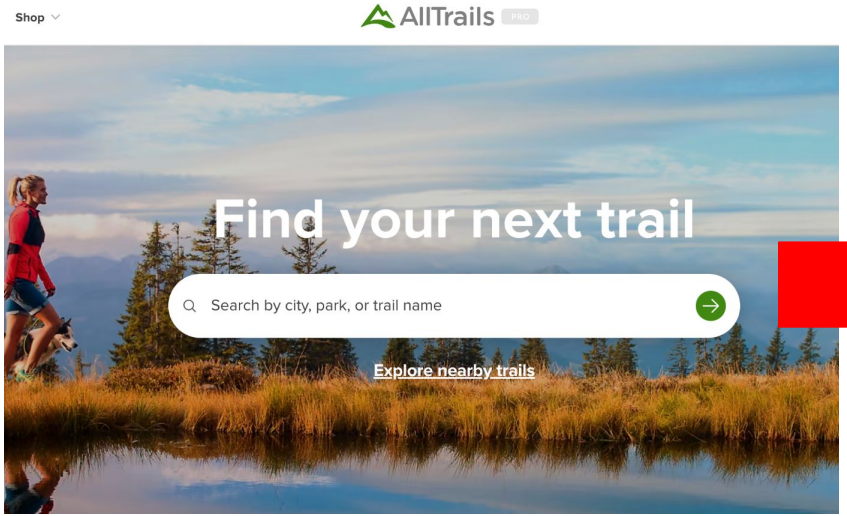
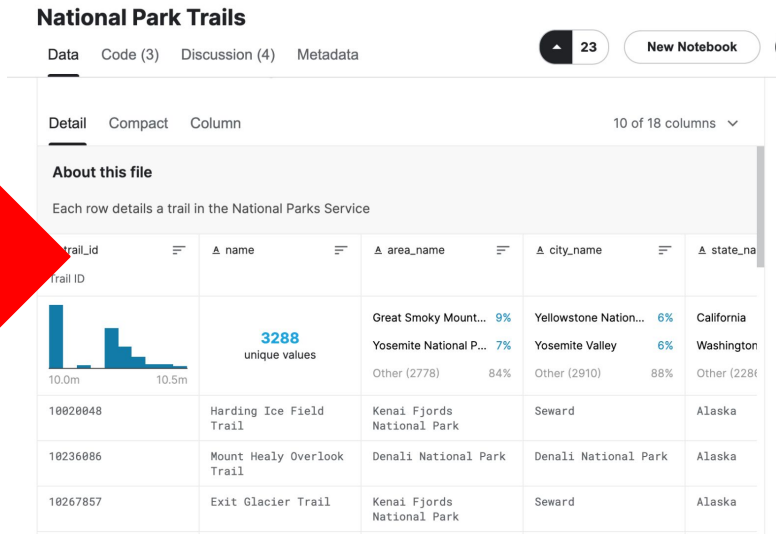
Trails in U.S. National Parks

Haley Johnson, uniqname: haleyej

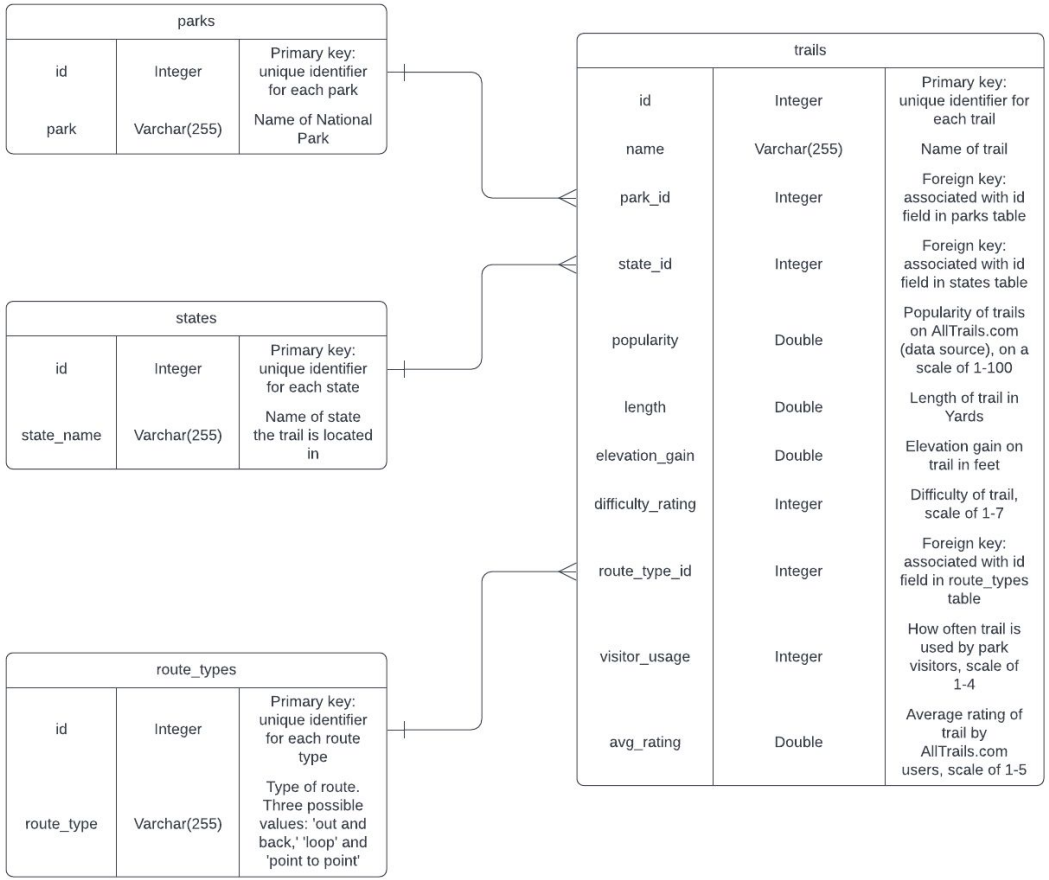
Data Source





uniqname: haleyej

Entity Relationship Diagram



uniqname: haleyej

Normalization

```python
def create_table(df, col):
    '''
    Turns column in the dataframe
    into a new dataframe that just
    contains the unique values in
    that column

    Function is used to split
    big dataframe into smaller table
    for normalization

    Returns a new dataframe, just
    made up of the target column
    '''
    temp = df[col].unique()
    df = pd.DataFrame(temp, columns = [col])
    df = df.reset_index()
    df = df.rename(columns = {'index': 'id'})
    df['id'] = df['id'].apply(lambda s: s + 1)
    return df

```

```python
def normalize(df1, df2, target, fk):
    '''
    Normalize tables to prepare for SQL
    puts foreign keys into main table

    Takes in four arguments:
    two dataframes that are being marged,
    the column used to merge them
    the foreign key connecting the table

    Return a dataframe with the
    foreign key normalized
    '''
    df1 = df1.merge(df2, on = target)
    df1 = df1.rename(columns = {'id': fk})
    df1 = df1.drop(columns = target)
    return df1
```

uniqname: haleyej

Homework Question

Give me a list of trails and the parks they're located in that have high difficulty ratings (greater than or equal to 5), but are frequently used by visitors.

- **Note:** difficulty is reported on a 1-7 scale, visitor usage is reported on a 1-4 score

uniqname: haleyej

Homework Question

Give me a list of trails and the parks they're located in that have high difficulty ratings (greater than or equal to 5), but are frequently used by visitors.

- **Note:** difficulty is reported on a 1-7 scale, visitor usage is reported on a 1-4 score

```
cur.execute('''SELECT t.name, p.park, t.difficulty_rating, t.visitor_usage
               FROM trails t
               JOIN parks p
               ON t.park_id = p.id
               WHERE t.difficulty_rating >= 5
               AND t.visitor_usage =
                   (SELECT max(visitor_usage) FROM trails)''')
```

uniqname: haleyej

Homework Question

Give me a list of trails and the parks they're located in that have high difficulty ratings (greater than or equal to 5), but are frequently used by visitors.

- **Note:** difficulty is reported on a 1-7 scale, visitor usage is reported on a 1-4 score

```
Merced Lake High Sierra Camp Trail in Yosemite National Park
Visitor Usage Rating (1-4): 4.0
Difficulty Rating: 5

Alum Cave Trail to Mount LeConte in Great Smoky Mountains National Park
Visitor Usage Rating (1-4): 4.0
Difficulty Rating: 5

Chimney Tops Trail in Great Smoky Mountains National Park
Visitor Usage Rating (1-4): 4.0
Difficulty Rating: 5

Angels Landing Trail in Zion National Park
Visitor Usage Rating (1-4): 4.0
Difficulty Rating: 5

Old Rag Mountain Loop in Shenandoah National Park
Visitor Usage Rating (1-4): 4.0
Difficulty Rating: 5

Lena Lake Trail to Upper Lena Lake Trail in Olympic National Park
Visitor Usage Rating (1-4): 4.0
Difficulty Rating: 5
```

uniqname: haleyej

Things I'd Do Differently

features	activities
['dogs-no', 'forest', '... 5%	['birding', 'hiking', 'n... 14%
['dogs-no', 'views', '... 4%	['birding', 'hiking', 'n... 8%
Other (3017) 91%	Other (2614) 79%
['dogs-no', 'forest', 'river', 'views', 'waterfall', 'wild-flowers', 'wildlife']	['birding', 'camping', 'hiking', 'nature-trips', 'trail-running']
['dogs-no', 'forest', 'views', 'wild-flowers', 'wildlife']	['birding', 'camping', 'hiking', 'nature-trips', 'walking']
['dogs-no',	['hiking',

Linking tables

i.e. activities
parks_to_activities

uniqname: haleyej