pdm | Projects | Branches | Notebook | Pdm | Pdm | Settings

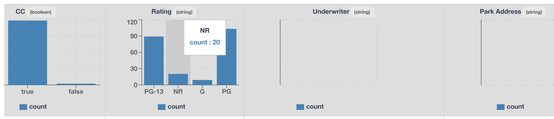
On branch **Untitled Branch**

[1]
```
LOAD DATASET pdm AS com.crealytics.spark.excel FROM pdm.xlsx @ artifact file 37
```

Console | Timing | Datasets | Charts

pdm (221 rows)

Views



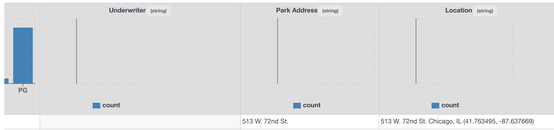
	PG-13		513 W. 72nd St.
	PG-13		5010 W. 50th St.
	PG-13		8930 S. Muskegon Ave.
	NR	The Cultural Services at the Consulate General of France in Chicago	5800 N. Lake Shore Dr.
	PG-13		515 S. Washtenaw Ave.
	PG-13		401 W. 123rd St.
	G		3309 S. Shields Ave.
	PG	The Wrightwood Neighbors Association	1001 W. Wrightwood Ave.
	PG		5800 N. Lake Shore Dr.
	G		2021 N. Burling St.

[2]
```
ds = vizierdb.get_dataset('pdm')
'''
issue-1: drop useless columns according to description that only have 1@ columns.
https: //data. cityofchicago.org/Events/Chicago-Park-District-Movies-in-the-Parks-2018/e2v8-k3us
columns_to_drop = ["Community Areas", "Boundaries - ZIP Codes", "Zip Codes", "Census Tracts", "Wards"]
'''
ds.delete_column( 'Community Areas')
ds.delete_column('Boundaries - ZIP Codes')
ds.delete_column('Zip Codes')
ds.delete_column( 'Census Tracts')
ds.delete_column( 'Wards')
vizierdb.update_dataset('pdm', ds)
```

Console | Timing | Datasets | Charts

pdm (221 rows)

Views



		513 W. 72nd St.	513 W. 72nd St. Chicago, IL (41.763495, -87.637669)
		5010 W. 50th St.	5010 W. 50th St. Chicago, IL (41.802024, -87.748205)
		8930 S. Muskegon Ave.	8930 S. Muskegon Ave. Chicago, IL (41.732921, -87.555452)
	The Cultural Services at the Consulate General of France in Chicago	5800 N. Lake Shore Dr.	5800 N. Lake Shore Dr. Chicago, IL
		515 S. Washtenaw Ave.	515 S. Washtenaw Ave. Chicago, IL (41.874341, -87.693699)
		401 W. 123rd St.	401 W. 123rd St. Chicago, IL (41.670629, -87.6324)
		3309 S. Shields Ave.	3309 S. Shields Ave. Chicago, IL (41.834212, -87.635226)
	The Wrightwood Neighbors Association	1001 W. Wrightwood Ave.	1001 W. Wrightwood Ave. Chicago, IL (41.929029, -87.653839
		5800 N. Lake Shore Dr.	5800 N. Lake Shore Dr. Chicago, IL
		2021 N. Burling St.	2021 N. Burling St. Chicago, IL (41.919044, -87.647303)

[3]
```
import re
ds = vizierdb.get_dataset('pdm')
'''
issue-2: Remove the duplicate information in the Location column and retain the latitude and longitude.
'''
def extract_coordinates(location_string):
    pattern = r'\(([-\d.]+, [-\d.]+)\)'
    match = re.search(pattern, location_string)
    if match:
        return match.group(0)
    else:
        return None

for row in ds.rows:
    loc = extract_coordinates(row.get_value('Location'))
    row.set_value('Location', loc)

vizierdb.update_dataset('pdm', ds)
```

Console | Timing | Datasets | Charts

pdm (221 rows)

Views

ne (string)	Title (string)	CC (boolean)	Rating (string)	Underwriter (string)	Park Address (string)	Location (string)
6174	Black Panther	true	PG-13		513 W. 72nd St.	(41.763495, -87.63766
6022	Justice League	true	PG-13		5010 W. 50th St.	(41.802024, -87.74820
6023	Star Wars: The Last Jedi	true	PG-13		8930 S. Muskegon Ave.	(41.732921, -87.55545
1134	Donkey Skin (Peau d'ane)	true	NR	The Cultural Services at the Consulate General of France in Chicago	5800 N. Lake Shore Dr.	
5001	Justice League	true	PG-13		515 S. Washtenaw Ave.	(41.874341, -87.69369
7090	The Fate of the Furious	true	PG-13		401 W. 123rd St.	(41.670629, -87.6324)
6012	Cars 3	true	G		3309 S. Shields Ave.	(41.834212, -87.63522
7816	Paddington 2	true	PG	The Wrightwood Neighbors Association	1001 W. Wrightwood Ave.	(41.929029, -87.65383
1134	9 to 5	true	PG		5800 N. Lake Shore Dr.	
7898	The Wizard of Oz	true	G		2021 N. Burling St.	(41.919044, -87.64730

[4]
```
import requests

ds = vizierdb.get_dataset('pdm')

"""
issue-3: Find the rows with missing latitude and longitude and
       use the Google API query address to populate the latitude and longitude
"""
api_key = 'AIzaSyDhSfxYmhNWFksZUWDFUVFJKrAApL9O7ZQ'

def get_coordinates_from_api(park_address):
    api_url = f'https://maps.googleapis.com/maps/api/geocode/json?address=\'{park_address}\'&key={api_key}'
    print(api_url)
    response = requests.get(api_url)
    data = response.json()
    if data['status'] == 'OK':
        location = data['results'][0]['geometry']['location']
        return f"({location['lat']}, {location['lng']})"
    else:
        return None

for row in ds.rows:
    if row.get_value('Location') is None:
        loc = get_coordinates_from_api(row.get_value('Park Address'))
        row.set_value('Location', loc)

vizierdb.update_dataset('pdm', ds)
```

Console | Timing | Datasets | Charts

pdm (221 rows)

Views

		513 W. 72nd St.	(41.763495, -87.637669)
		5010 W. 50th St.	(41.802024, -87.748205)
		8930 S. Muskegon Ave.	(41.732921, -87.555452)
	The Cultural Services at the Consulate General of France in Chicago	5800 N. Lake Shore Dr.	(41.9861626, -87.6529082)
		515 S. Washtenaw Ave.	(41.874341, -87.693699)
		401 W. 123rd St.	(41.670629, -87.6324)
		3309 S. Shields Ave.	(41.834212, -87.635226)
	The Wrightwood Neighbors Association	1001 W. Wrightwood Ave.	(41.929029, -87.653839)
		5800 N. Lake Shore Dr.	(41.9861626, -87.6529082)
		2021 N. Burling St.	(41.919044, -87.647303)

[5]
```
from datetime import datetime
"""
issue-4: Verification time and day of the week
"""
ds = vizierdb.get_dataset('pdm')

match_all = True

for row in ds.rows:
    date = datetime.strptime(row.get_value('Date'), '%m/%d/%Y').strftime('%A')
    day_of_week = date[:3]
    if day_of_week != row.get_value('Day'):
        match_all = False
        print(f"Day and Date mismatch in row {row.get_value('Day')} vs {date}")

if match_all:
    print("All rows match")

"""
issue-5: Verification Location
"""
all_locations_not_none = True

for row in ds.rows:
    location = row.get_value('Location')
    if location is None:
        all_locations_not_none = False
        print(f"None value found in Location column in row {row}")

if all_locations_not_none:
    print('All rows have non-None values in the Location column")

"""
This is because api(https://maps.googleapis.com/maps/api/geocode/json?address='8050 S. Chapel.'&key={api_key})
return ZERO_RESULTS.
"""
```

Console | Timing | Datasets | Charts

```
n row <Wed, 07/18/2018, Rosenblum Park, (312) 747-6649, Jumanji: Welcome to the Jungle, True, PG-13, None, 8050 S. Chapel, None>
```

+