```
In [40]: from IPython.display import Image
         from IPython.display import display, Image
         display(Image(filename='Morning Fuel.png'))
```

Recommendation

we recommend Morning Fuel courses on this neighborhoods, they meet the neccesary criterias;

- Central Park,
- Five Points,
- Gateway-Green Valley Ranch,
- Highland,
- University

CONLUSIONS

AURARIA AND CBD ARE THE NEIGHBORHOODS WITH THE HIGHEST PEOPLE AGE OF 18-35, BUT THEY DO NOT HAVE AFFLUENT HOUSEHOLDS. LEAVING THOSE TWO OUT (which can still be used later on in contemplating other neighborhoods to open new stores), UNIVERSITY HAS THE HIGHEST PEOPLE AGE OF 18-35 YEARS. WE GET THE TOP 3 NEIGHBORHOODS BY APPLYING THE PERCENTAGE OF AFFLUENT HOUSEHOLDS TO THE AMOUNT OF 18-35 YEAR OLD PEOPLE (PEOPLE WHO WILL ACTUALLY BUY).

FIVE POINTS-HIGHLAND-UNIVERSITY

Introduction

Polarity Ventures is assisting a customer, Morning Fuel, who runs coffee shops throughout Colorado. The coffee shops of the company sell high-quality, responsibly sourced coffee, pastries, and sandwiches. They have three Fort Collins locations and plan to expand into Denver.

Our customer believes that a new store should be located near rich households, and the store caters to the 20-35 year old population.

To help with the search, the Analytic team gathered geographical and demographic information on Denver's neighborhoods. We also gathered information from Starbucks locations in Denver. Starbucks and Morning Fuel do not competing for the same customers; the team used their location as an example.

For the purpose of this analysis, the end results will be providing Morning Fuel with the best neighborhood locations in denver where the new shops should be located. By providing Morning Fuel with the following information;

- A visualization of Denver's neighborhoods and the Starbucks store locations.
- Find the neighborhoods with the highest proportion of people in the target demographic.
- Select the top three neighborhoods where your client should focus their search.

Data

Starbucks locations were scrapped from the Starbucks store locator webpage by Chris Meller. Statistical Neighborhood information from the City of Denver Open Data Catalog, CC BY 3.0 license. Census information from the United States Census Bureau. Publicly available information.

Data Description

You have assembled information from three different sources (locations, neighborhoods, demographics):

Starbucks locations in Denver, Colorado

- "StoreNumber" - Store Number as assigned by Starbucks
- "Name" - Name identifier for the store
- "PhoneNumber" - Phone number for the store
- "Street 1, 2, and 3" - Address for the store
- "PostalCode" - Zip code of the store
- "Longitude, Latitude" - Coordinates of the store

Neighborhoods' geographical information

- "NBHD_ID" - Neighborhood ID (matches the census information)
- "NBHD_NAME" - Name of the statistical neighborhood
- "Geometry" - Polygon that defines the neighborhood

Demographic information

- "NBHD_ID" - Neighborhood ID (matches the geographical information)
- "NBHD_NAME' - Nieghborhood name
- "POPULATION_2010' - Population in 2010
- "AGE_ " - Number of people in each age bracket (< 18, 18-34, 35-65, and > 65)
- "NUM_HOUSEHOLDS" - Number of households in the neighborhood
- "FAMILIES" - Number of families in the neighborhood
- "NUM_HHLD_100K+" - Number of households with income above 100 thousand USD per year

In [1]:
```
%%capture
pip install geopandas
```

In [2]:
```
# Importing pandas
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

# Reading in the data
census = pd.read_csv('census.csv')
denver = pd.read_csv('denver.csv')
neighborhoods = gpd.read_file("C:/Users/hp/Desktop/Projects/Coffee Shop/neighborhoods.shp")
neighborhoods
```

Out[2]:

	NBHD_ID	NBHD_NAME	geometry
0	2	Auraria	POLYGON ((-105.00042 39.74552, -105.00041 39.7...
1	21	Cory - Merrill	POLYGON ((-104.94070 39.69540, -104.94070 39.6...
2	7	Belcaro	POLYGON ((-104.94070 39.71156, -104.94069 39.7...
3	70	Washington Park	POLYGON ((-104.95931 39.71566, -104.95931 39.7...
4	71	Washington Park West	POLYGON ((-104.97342 39.68982, -104.97356 39.6...
...	...	...	...
73	77	Whittier	POLYGON ((-104.95977 39.75072, -104.96038 39.7...
74	18	Cole	POLYGON ((-104.95975 39.76199, -104.96037 39.7...
75	76	Westwood	POLYGON ((-105.03970 39.71125, -105.03849 39.7...
76	62	Sunnyside	POLYGON ((-104.99818 39.78256, -104.99820 39.7...
77	23	DIA	POLYGON ((-104.76269 39.79833, -104.76290 39.7...

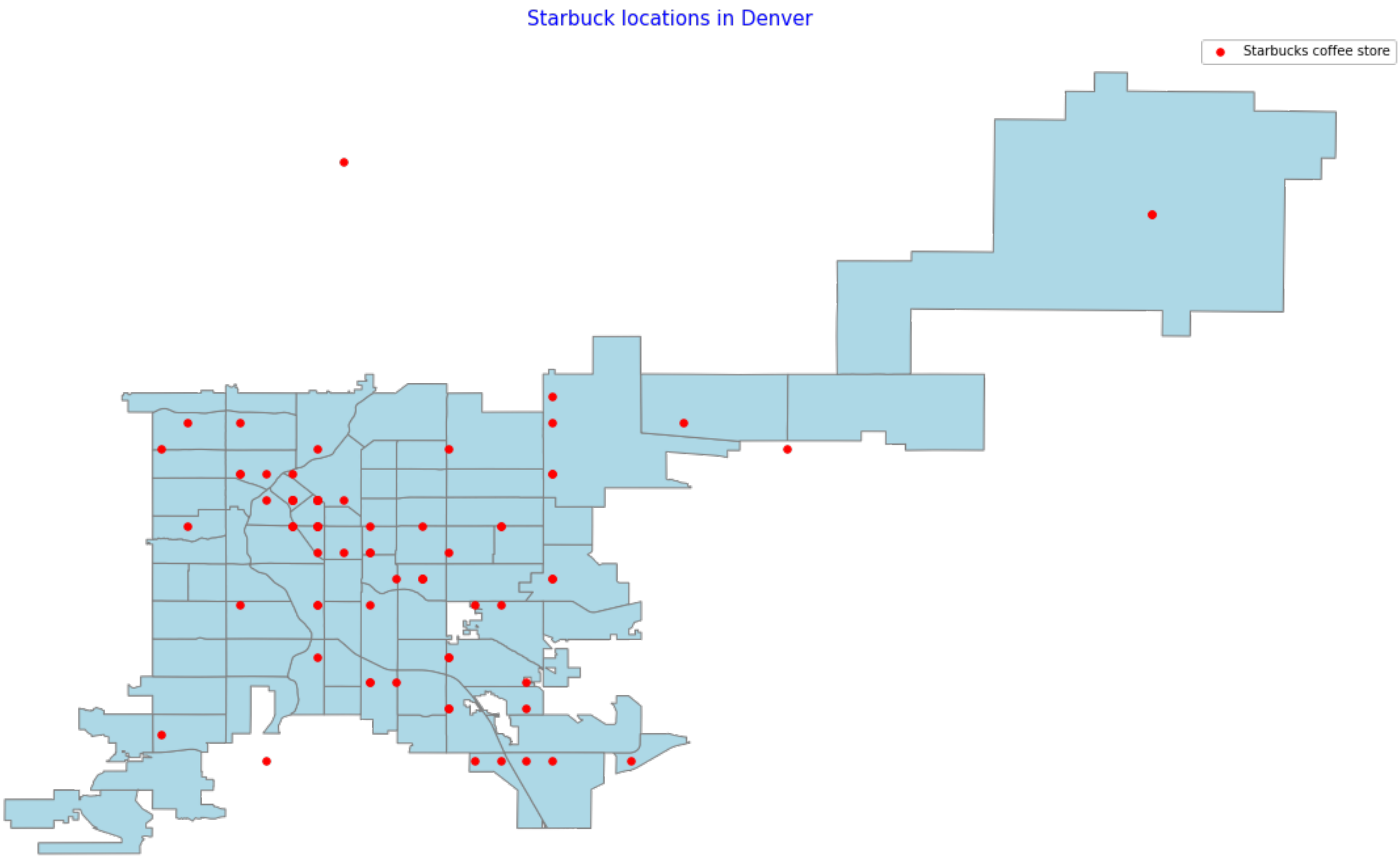
78 rows × 3 columns

A visualization of Denver's Neighborhoods and the Starbucks store locations.

Since we'll be using Starbucks data as a landmark, let's have a look at where each Starbucks location is. By knowing this and other information when we dig further into the data, we'll be able to determine crucial spots where additional stores should be opened.

In [3]:
```
#convert denver data to geographical data
denver_geo = gpd.GeoDataFrame(denver, geometry=gpd.points_from_xy(denver.Longitude, denver.Latitude))


fig, ax = plt.subplots(figsize=(20,20))
ax.set_aspect('equal')
ax=neighborhoods.plot(ax=ax,color='lightblue', edgecolor='grey', legend =True)
ax.set_axis_off()
denver_geo.plot(ax=ax, markersize = 30, color = 'red', marker = 'o', label='Starbucks coffee store')
plt.title('Starbuck locations in Denver', color='blue',size=15)
plt.legend()
plt.show();
```



A map of Denver neighborhoods showing where starbucks are located.

Find the neighborhood with the highest proportion of people in the target demographic

To get the proportion of target customer we will be creating columns to hold this information.

We create some columns analyzing the desired variables.
ageprop_18_34 is the target age percentage
prop_highincome is a percentage of the amount of affluent hh
ind is index where we apply the high income households % to the amount of 18-35 age people. With this we can approach the amount of real target people in the neighborhood.

NOTE: Like the desired age rate (20-35) is not in the census as that, we will focus in 18-34 age rate that contains the desired one as it is prety similar.

In [4]:
```
# proportion of people betwenn 18 - 34:
census['ageprop_18_34'] = census['AGE_18_TO_34'] / census['POPULATION_2010']
#for the missing figures in the family column fill wih the mean
over_100k_mean = census[census['NUM_HHLD_100K+'].notna()]['NUM_HHLD_100K+'].sum() / census[census['NUM_HHLD_100K+'].notna()]['NUM_H
# proportion of housholds with high income:
census['prop_highincome'] = census['NUM_HHLD_100K+'] / census['NUM_HOUSEHOLDS']
#  real target customers in the neighborhood.
census['ind']=census['prop_highincome']*census['AGE_18_TO_34']
```

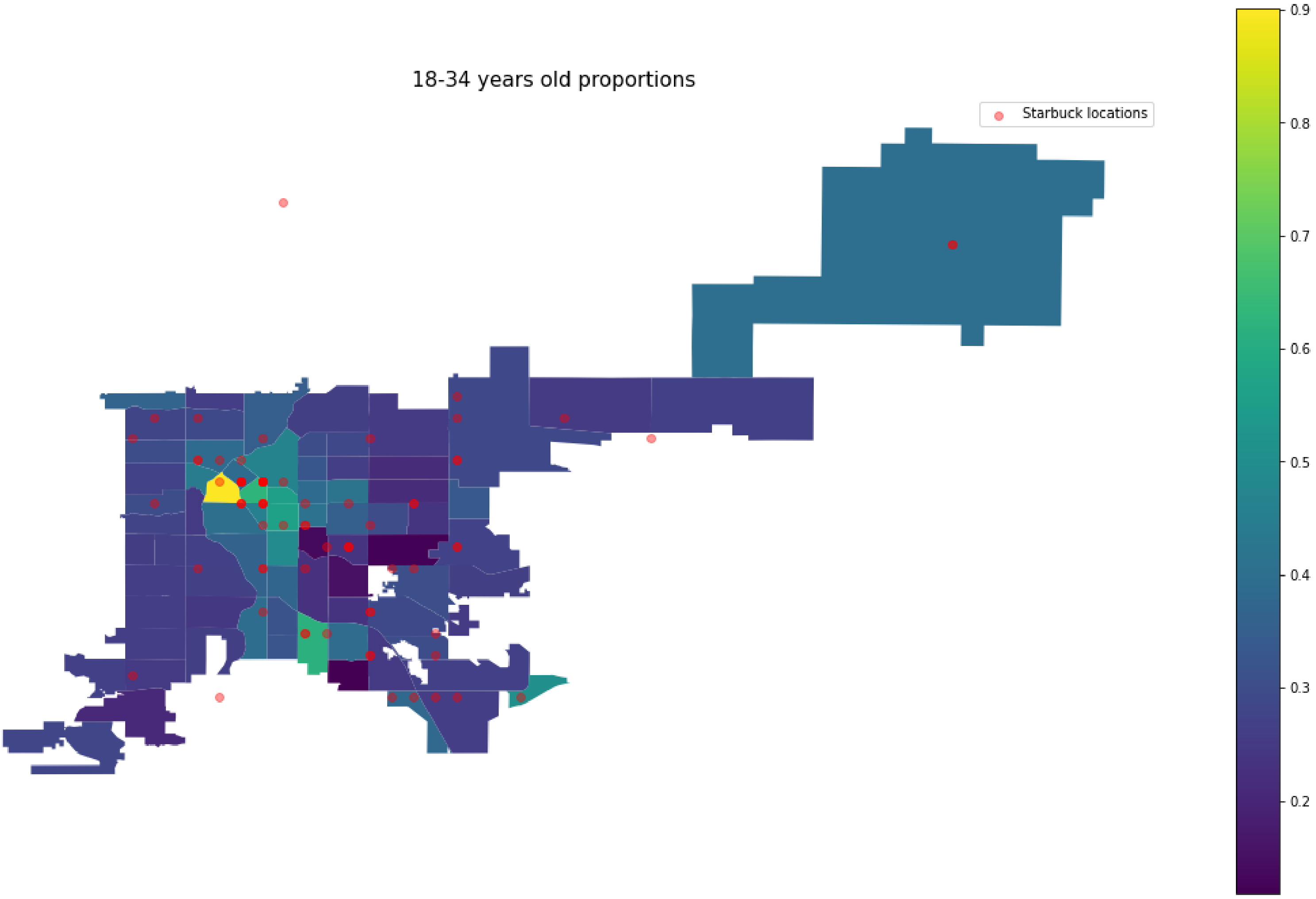
In [12]:
```
# Calculate average proportion of people of age 18-34:
avg_prop_age_18_34 = census['ageprop_18_34'].mean()
#avg_prop_age_18_34
```

```
neighborhood_census = neighborhoods.merge(census, on='NBHD_ID', how='left').fillna(0)
neighborhoods

# Renaming columns
neighborhood_census.rename(columns={"NBHD_NAME_x":"NBHD_NAME","ind_y": "ind","POPULATION_2010_x":"POPULATION_2010","AGE_18_

#Deleting duplicate column
for col in neighborhood_census.columns:
    if 'NBHD_NAME_y' in col:
        del neighborhood_census[col]
fig, ax1 = plt.subplots(figsize=(20,12))

# A heat map showing Age Demographic
neighborhood_census.plot(column='ageprop_18_34',legend=True,ax=ax1, label ='Proportion of people age 18 to 34')
ax1.set_axis_off()
denver_geo.plot(ax=ax1, color='red',legend=True,alpha=0.4, label='Starbuck locations')
ax1.legend()
ax1.set_title('18-34 years old proportions',size=15);
```

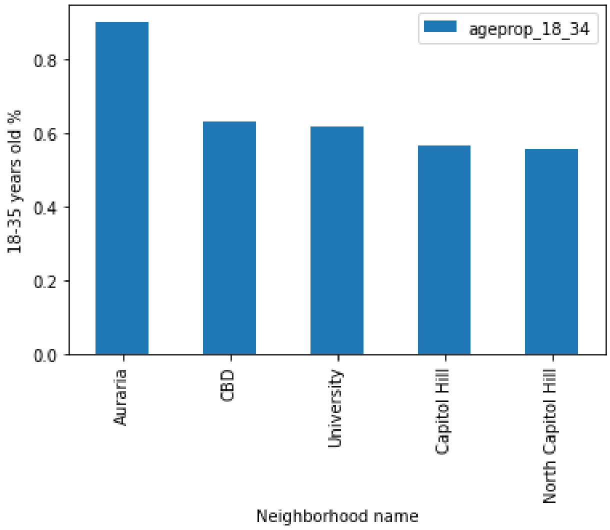


From the above map, using the scale to the proportion of people from age 18 to 34 with the highest proportion in yellow and less in purple.

Target Neighborhoods

Now let's begin to narrow down our analysis, to get the neighborhoods that Morning Fuel will targeting in Denver.

In [13]:
```
# Target Neigborhoods
target_age_per=neighborhood_census.sort_values('ageprop_18_34', ascending=False).head()
target_age_per.plot(x='NBHD_NAME', y='ageprop_18_34', kind='bar', ylabel='18-35 years old %', xlabel='Neighborhood name');
```

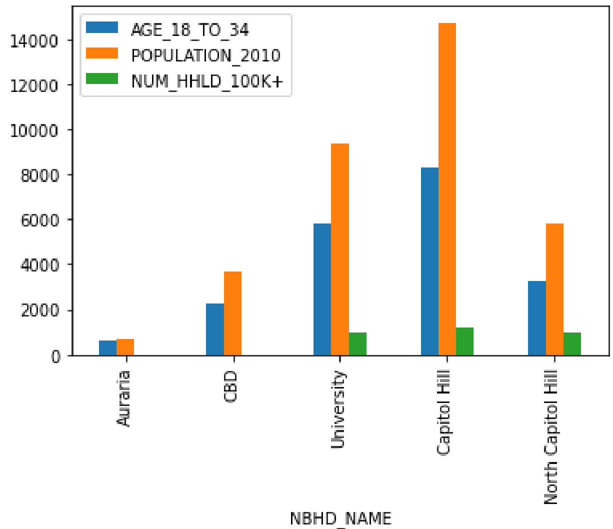


Neighborhoods with highest proportion of 18 to 34 are;

- Auraria
- CBD
- University
- Capitol Hill
- North Capitol Hill

Now to see if this neighborhoods meet our other criterias

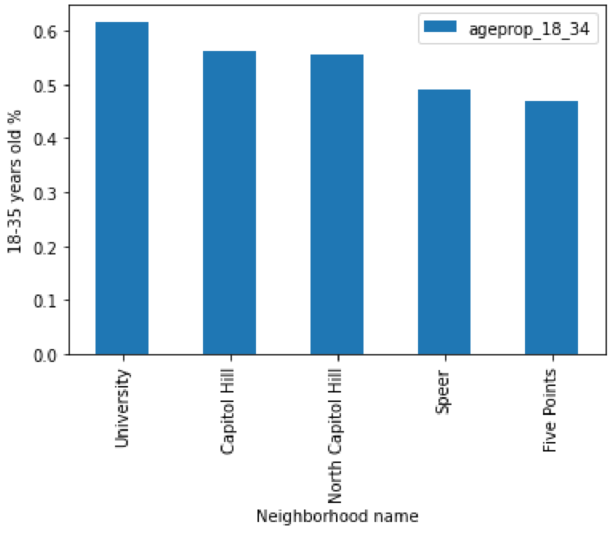
In [14]:
```
# Plotting all critrias each neighborhood has to meet to be a target market
target_age_per.head().plot(x='NBHD_NAME',y=['AGE_18_TO_34','POPULATION_2010','NUM_HHLD_100K+'], kind='bar');
```



From this plots, we can see that the 2 neighbourhoods have the target age demographic **Auraria** and **CBD**, but they lack the second critria needeed, **Household making above 100k+**

In [28]:
```
# Removing all neighborhoods that dont meet the target market
census_aff=neighborhood_census[neighborhood_census['NUM_HHLD_100K+']>0]
census_aff.sort_values('ageprop_18_34', ascending=False)

# A visual Neighborhoods meeting the critria
census_aff.sort_values('ageprop_18_34',ascending=False).head().plot(x='NBHD_NAME',y='ageprop_18_34',kind='bar', ylabel='18-35 years old %', xl
```



After removing Auraria and CBD, we can see that top 5 neighborhoods with most 18-35 % and that have affluent households.
1) University
2) Capitol Hill
3) North Capitol Hill
4) Speer
*5) Five Points

Using Statistic analysis, getting the statistic distribution of our census data we can get a better understanding of the neigborhood

In [16]:
```
census_aff.describe()
```

Out[16]:

	NBHD_ID	POPULATION_2010	AGE_LESS_18	AGE_18_TO_34	AGE_35_TO_65	AGE_65_PLUS	NUM_HOUSEHOLDS	FAMILIES	NUM_HHLD_
count	48.000000	48.000000	48.000000	48.000000	48.000000	48.000000	48.000000	48.000000	48.0
mean	42.604167	10273.854167	2260.937500	3133.312500	3814.270833	1065.333333	4456.187500	2147.625000	1095.2
std	22.169309	5006.996492	2022.145116	1674.986518	1708.502290	634.308749	2030.361562	1202.393281	650.8
min	1.000000	5327.000000	194.000000	994.000000	1867.000000	360.000000	1695.000000	461.000000	120.0
25%	26.750000	7427.750000	1271.750000	2007.750000	2881.250000	706.250000	3321.500000	1395.000000	439.2
50%	42.500000	8823.500000	2015.500000	2692.000000	3453.000000	931.500000	3840.000000	1896.500000	1043.5
75%	62.500000	11756.750000	2606.000000	3642.750000	4329.250000	1229.250000	5240.250000	2446.000000	1680.0
max	78.000000	30348.000000	11137.000000	8274.000000	10405.000000	3383.000000	10856.000000	7056.000000	2748.0

Applying the statistical analysis results into the data to pin point the target markets

We would like a neighborhood with a good population and affluent households aside a good percentage of target age (20-35). We can see 75% of pop (11756), 18-34 age (3642.75) and NUM_HHLD_100K+ (1680)

In [21]:
```
denver_best = neighborhood_census[(neighborhood_census['POPULATION_2010']>11756.75) | (neighborhood_census['AGE_18_TO_34']>3642.75)
denver_best.drop(columns='index', inplace=True)
final=denver_best[['NBHD_NAME','geometry','POPULATION_2010','AGE_18_TO_34','ageprop_18_34','prop_highincome','ind']].sort_values(['ageprop
final.head()
```

Out[21]:

	NBHD_NAME	geometry	POPULATION_2010	AGE_18_TO_34	ageprop_18_34	prop_highincome	ind
5	University	POLYGON ((-104.95945 39.68473, -104.95945 39.6...	9375.0	5784.0	0.616960	0.263634	1524.858739
11	Capitol Hill	POLYGON ((-104.97341 39.74003, -104.97299 39.7...	14708.0	8274.0	0.562551	0.112657	932.120671
1	Speer	POLYGON ((-104.97325 39.71840, -104.97323 39.7...	10954.0	5361.0	0.489410	0.230285	1234.556681
15	Five Points	POLYGON ((-104.97192 39.77030, -104.97335 39.7...	12712.0	5961.0	0.468927	0.262254	1563.296909
4	Highland	POLYGON ((-104.99820 39.76930, -104.99821 39.7...	8429.0	3269.0	0.387828	0.412491	1348.432825

In [29]:
```
# Sorting the data to show the top neighborhoods by age proportion of target demo and affluent household proportion
denver_best.sort_values(['ageprop_18_34','prop_highincome'], ascending=([False,False])).head(5)
```

Out[29]:

	NBHD_ID	NBHD_NAME	geometry	POPULATION_2010	AGE_LESS_18	AGE_18_TO_34	AGE_35_TO_65	AGE_65_PLUS	NUM_HOUSEHOLDS	FAI
5	64	University	POLYGON ((-104.95945 39.68473, -104.95945 39.6...	9375.0	826.0	5784.0	2246.0	519.0	3759.0	
11	9	Capitol Hill	POLYGON ((-104.97341 39.74003, -104.97299 39.7...	14708.0	408.0	8274.0	5155.0	871.0	10856.0	
1	59	Speer	POLYGON ((-104.97325 39.71840, -104.97323 39.7...	10954.0	707.0	5361.0	3970.0	916.0	7304.0	
15	26	Five Points	POLYGON ((-104.97192 39.77030, -104.97335 39.7...	12712.0	1446.0	5961.0	4660.0	645.0	6406.0	
4	36	Highland	POLYGON ((-104.99820 39.76930, -104.99821 39.7...	8429.0	1296.0	3269.0	3070.0	794.0	4131.0	

In [30]:

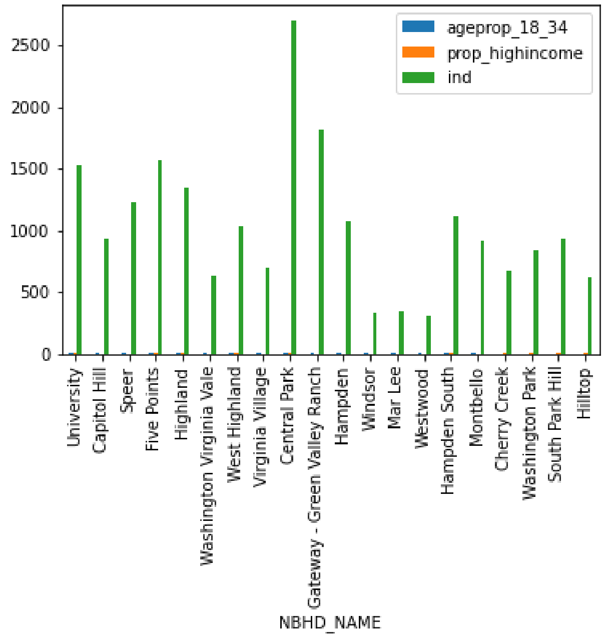
```
# Sorting the data to show the top neighborhoods by the real target people in the neighborhood.
denver_best.sort_values('ind', ascending=False).head(5)
```

Out[30]:

	NBHD_ID	NBHD_NAME	geometry	POPULATION_2010	AGE_LESS_18	AGE_18_TO_34	AGE_35_TO_65	AGE_65_PLUS	NUM_HOUSEHOLDS	FAI
12	60	Central Park	POLYGON ((-104.86604 39.79841, -104.86604 39.7...	13948.0	3516.0	4008.0	6045.0	379.0	4092.0	
14	28	Gateway - Green Valley Ranch	POLYGON ((-104.80990 39.77283, -104.80988 39.7...	29201.0	10074.0	7778.0	10405.0	944.0	9427.0	
15	26	Five Points	POLYGON ((-104.97192 39.77030, -104.97335 39.7...	12712.0	1446.0	5961.0	4660.0	645.0	6406.0	
5	64	University	POLYGON ((-104.95945 39.68473, -104.95945 39.6...	9375.0	826.0	5784.0	2246.0	519.0	3759.0	
4	36	Highland	POLYGON ((-104.99820 39.76930, -104.99821 39.7...	8429.0	1296.0	3269.0	3070.0	794.0	4131.0	

In [25]:

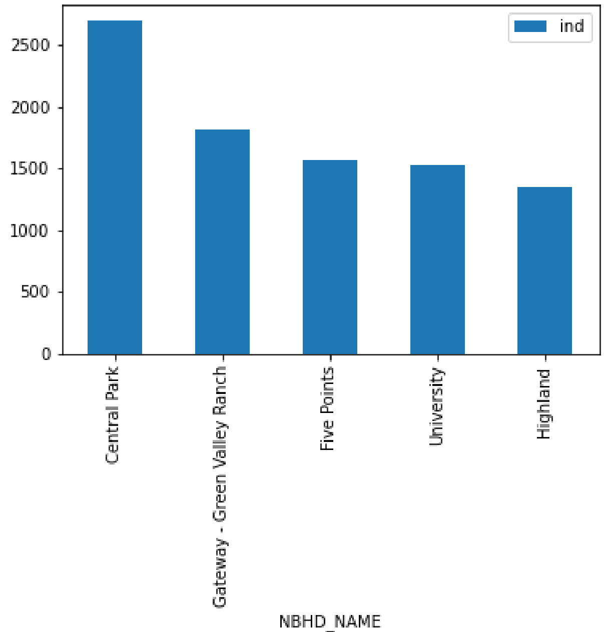
```
# The top neighborhoods by the real target people in the neighborhood.
final.plot(x='NBHD_NAME', y=['ageprop_18_34','prop_highincome','ind'], kind='bar');
```



After applying the 75% filter, we get the five best neighborhoods (regarding target age %) for a new coffee shop are:
1) University
2) Capitol Hill
3) Speer Hill
4) Five Points
*5) Highland

North Capitol stayed behind after last "75%" filter

In [26]:
```
top_three=final.sort_values('ind', ascending=False).head()
top_three.plot(x='NBHD_NAME', y='ind', kind='bar');
```



If we sort the top five using the index (% of Affluent HH * amount 18-35 people), we see that Five Point is the neighborhood were we have a total population, target age population and Affluent HH over 75% of their values and also % of Affleunt HH influence over amount of 18-35 age people higher than others.

In [27]:
```
fig, ax = plt.subplots(figsize=(8,8))
ax=top_three.plot(ax=ax, column='NBHD_NAME', legend=True)
ax.set_axis_off()
```