labor_market

March 16, 2025

1 Labor Market Projection Data

[22]:
```
# Data science libraries
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

# Local modules
import scripts.cccco as cccco
import scripts.labor_market as labor_market
import scripts.join_tools as join_tools
```

[2]:
```
all_colleges = cccco.get_ccc_colleges()
```

The issue here is that labor market data is indexed by **Standard Occupation Classification** (SOC) Codes, while college programs are indexed by **Taxonomy of Program** (TOP) Codes and by **Classification of Instructional Programs** (CIP) Codes. The match between these is not one-to-one.

TOP Codes do not appear to be listed in .csv or .xlsx form anywhere I can think to look, but they are indirectly present in the TOP-CIP crosswalk (`./resources/TOPCIPJune26.xlsx`). Active TOP Codes can also be mined from DataMart.

Annoyingly, there is not crosswalk between TOP Codes and SOC Codes, so we'll have to map from TOP to CIP to SOC. Alternatively, we can try fuzzy joins between TOP Code Descriptions (scraped from DataMart) and SOC Code Descriptions (`./resources/soc_2018_definitions.xlsx`).

First, add the Metropolitan Statistical Area to which the college belongs to the college dataset:

[3]:
```
# Gather all Metropolitan Statistical Area (MSA) data. This will allow us to
# join labor market data to the college data.

msa = labor_market.get_labor_market_msa_data()
print(f"{msa.shape[0]:,} rows")
msa.head()
```

```
29 rows
```

[3]:
```
                                        Area Name  \
0                      Redding MSA (Shasta County)
```

```
1  Oakland-Hayward-Berkeley MD (Alameda and Contr…
2                        Fresno MSA (Fresno County)
3  San Jose-Sunnyvale-Santa Clara MSA (San Benito…
4  Riverside-San Bernardino-Ontario MSA (Riversid…

          Metropolitan Statistical Area                          MSA Counties
0                         Redding MSA                          Shasta County
1           Oakland-Hayward-Berkeley MD    Alameda and Contra Costa Counties
2                          Fresno MSA                          Fresno County
3    San Jose-Sunnyvale-Santa Clara MSA   San Benito and Santa Clara Counties
4  Riverside-San Bernardino-Ontario MSA  Riverside and San Bernardino Counties
```

[4]:
```
# Add MSA data to `all_colleges` dataset
area_counties = msa["MSA Counties"].dropna().unique()

all_colleges["Matched Counties"] = all_colleges["County"].apply(
    lambda x: join_tools.get_best_match(x, area_counties)
)

# One row per college per program
all_colleges = pd.merge(
    all_colleges,
    msa,
    left_on="Matched Counties",
    right_on="MSA Counties",
    how="left"
)

print(f"{all_colleges.shape[0]:,} rows")
all_colleges.head()
```

```
115 rows
```

[4]:
```
  CollegeID              CollegeName DistrictID                StreetAddress  \
0       021         Cuyamaca College        020  900 Rancho San Diego Parkway
1       022        Grossmont College        020  8800 Grossmont College Drive
2       031  Imperial Valley College        030            380 East Aten Road
3       051        MiraCosta College        050               1 Barnard Drive
4       061          Palomar College        060        1140 West Mission Road

         City      County    Zip ZipPlus4                MailingAddress  \
0    El Cajon  San Diego   92019     4304  900 Rancho San Diego Parkway
1    El Cajon  San Diego   92020     1799  8800 Grossmont College Drive
2    Imperial   Imperial   92251     9787            380 East Aten Road
3   Oceanside  San Diego   92056     3899               1 Barnard Drive
4  San Marcos  San Diego   92069     1487        1140 West Mission Road

  MailingCity  …          Phone         WebsiteURL   Latitude   Longitude  \
```

```
0    El Cajon  …  619.660.4000    www.cuyamaca.edu  32.744890 -116.935229
1    El Cajon  …  619.644.7000   www.grossmont.edu  32.817897 -117.005640
2    Imperial  …  760.352.8320    www.imperial.edu  32.825859 -115.502999
3   Oceanside  …  760.757.2121   www.miracosta.edu  33.188864 -117.301064
4  San Marcos  …  760.744.1150     www.palomar.edu  33.147015 -117.183980

                            LogoURL District  Matched Counties  \
0           CuyamacaCollegeLogo.jpg     None  San Diego County
1          GrossmontCollegelogo.jpg     None  San Diego County
2  ImperialValleyCollegeLogocopy.jpg     None   Imperial County
3     Mira_Costa_College_Logo_4c.png     None  San Diego County
4             PalomarCollegeLogo.jpg     None  San Diego County

                                  Area Name Metropolitan Statistical Area  \
0  San Diego-Carlsbad MSA (San Diego County)        San Diego-Carlsbad MSA
1  San Diego-Carlsbad MSA (San Diego County)        San Diego-Carlsbad MSA
2            El Centro MSA (Imperial County)                El Centro MSA
3  San Diego-Carlsbad MSA (San Diego County)        San Diego-Carlsbad MSA
4  San Diego-Carlsbad MSA (San Diego County)        San Diego-Carlsbad MSA

       MSA Counties
0  San Diego County
1  San Diego County
2   Imperial County
3  San Diego County
4  San Diego County

[5 rows x 21 columns]
```

Next, we can join relevant labor market data by matching the TOP Code from the college data set to the SOC Code from the labor market data set. Since this is a one-to-many match, we won't do this at the DataFrame level.

Instead, we can perhaps consider the following use case:

1. User selects a job of interest (e.g., "Data Science").
2. We locate the relevant SOC Code data from labor market data set.
3. Rank MSA's by growth in that field (by percent growth or numeric?)
4. Report the colleges in the MSA with the top growth and add programs with corresponding TOP Codes.

[5]:
```python
# Step 1, 2, and 3:

occupation = "Computer"
occupation_by_msa = labor_market.get_occupation_projections_by_title(occupation)
print(f"{occupation_by_msa.shape[0]:,} rows")
occupation_by_msa.head()
```

```
320 rows
```

[5]:

	_id	_full_text	\
0	2040	'-2030':8 '-3021':11 '11':10 '110':17 '136088'…	
1	3804	'-2030':10 '-3021':13 '10':21,23 '106152':27 '…	
2	5646	'-2030':8 '-3021':11 '11':10 '120':22 '136026'…	
3	15694	'-2030':10 '-3021':13 '11':12 '120':24 '155953…	
4	11604	'-2030':20 '-3021':23 '102.84':36 '11':22 '125…	

	Area Type	Area Name	\
0	Metropolitan Area	Chico MSA (Butte County)	
1	Metropolitan Area	Hanford-Corcoran MSA (Kings County)	
2	Metropolitan Area	Modesto MSA (Stanislaus County)	
3	Metropolitan Area	Vallejo-Fairfield MSA (Solano County)	
4	Metropolitan Area	San Francisco-Redwood City-South San Francisco…	

	Period	SOC Level	Standard Occupational Classification (SOC)	\
0	2020-2030	4	11-3021	
1	2020-2030	4	11-3021	
2	2020-2030	4	11-3021	
3	2020-2030	4	11-3021	
4	2020-2030	4	11-3021	

	Occupational Title	Base Year Employment Estimate	\
0	Computer and Information Systems Managers	110.0	
1	Computer and Information Systems Managers	30.0	
2	Computer and Information Systems Managers	180.0	
3	Computer and Information Systems Managers	190.0	
4	Computer and Information Systems Managers	19660.0	

	Projected Year Employment Estimate	Numeric Change	Percentage Change	\
0	160.0	50.0	45.5	
1	40.0	10.0	33.3	
2	230.0	50.0	27.8	
3	240.0	50.0	26.3	
4	24660.0	5000.0	25.4	

	Exits	Transfers	Total Job Openings	Median Hourly Wage	\
0	20.0	80.0	150.0	65.42	
1	10.0	20.0	40.0	51.04	
2	30.0	120.0	200.0	65.40	
3	40.0	120.0	210.0	74.98	
4	3600.0	12590.0	21190.0	102.84	

	Median Annual Wage	Entry Level Education	Work Experience	Job Training
0	136088.0	Bachelor's degree	5 years or more	None
1	106152.0	Bachelor's degree	5 years or more	None
2	136026.0	Bachelor's degree	5 years or more	None
3	155953.0	Bachelor's degree	5 years or more	None

```
4            213890.0     Bachelor's degree  5 years or more         None
```

[6]:
```
# Get the MSA with the highest employment percent change for each occupation

max_msa_for_occupation = (
    occupation_by_msa.groupby("Standard Occupational Classification (SOC)")
    .first()
    .reset_index()
)

max_msa_for_occupation
```

[6]:

	Standard Occupational Classification (SOC)	_id	_full_text	Area Type
0	11-3021	2040	'-2030':8 '-3021':11 '11':10 '110':17 '136088'…	Metropolitan Area
1	15-1211	3838	'-1211':13 '-2030':10 '10':19,21 '15':12 '20':…	Metropolitan Area
2	15-1221	11659	'-1221':23 '-2030':20 '1260':34 '15':22 '2000'…	Metropolitan Area
3	15-1231	2083	'-1231':11 '-2030':8 '10':20 '15':10 '20':18 '…	Metropolitan Area
4	15-1232	5689	'-1232':11 '-2030':8 '15':10 '170':21 '2020':7…	Metropolitan Area
5	15-1241	10504	'-1241':11 '-2030':8 '10':17,19 '101258':23 '1…	Metropolitan Area
6	15-1244	2887	'-1244':12 '-2030':9 '10':20,22 '100025':26 '1…	Metropolitan Area
7	15-1251	11665	'-1251':23 '-2030':20 '124906':34 '15':22 '158…	Metropolitan Area
8	15-1299	2090	'-1299':11 '-2030':8 '100':21 '15':10 '180':16…	Metropolitan Area
9	17-2061	11683	'-2030':20 '-2061':23 '1380':32 '17':22 '19868…	Metropolitan Area
10	17-2072	3287	'-2030':8 '-2072':11 '100':16 '103371':24 '110…	Metropolitan Area
11	25-1021	9273	'-1021':19 '-2030':16 '0.00':31 '126702':32 '1…	Metropolitan Area
12	43-9071	1073	'-2030':8 '-9071':11 '0':19 '0.0':20 '18.86':2…	Metropolitan Area
13	49-2011	2382	'-2011':11 '-2030':8 '18.62':26 '20':21,23 '20…	Metropolitan Area
14	51-9161	14136	'-2030':12 '-9161':15 '10':23 '16.7':24 '20':2…	Metropolitan Area
15	51-9162	8089	'-2030':14 '-9162':17 '180':28 '2020':13 '210'…	Metropolitan Area

	Area Name	Period	SOC Level	\
0	Chico MSA (Butte County)	2020-2030	4	
1	Hanford-Corcoran MSA (Kings County)	2020-2030	4	
2	San Francisco-Redwood City-South San Francisco…	2020-2030	4	
3	Chico MSA (Butte County)	2020-2030	4	
4	Modesto MSA (Stanislaus County)	2020-2030	4	
5	Salinas MSA (Monterey County)	2020-2030	4	
6	El Centro MSA (Imperial County)	2020-2030	4	
7	San Francisco-Redwood City-South San Francisco…	2020-2030	4	
8	Chico MSA (Butte County)	2020-2030	4	
9	San Francisco-Redwood City-South San Francisco…	2020-2030	4	
10	Fresno MSA (Fresno County)	2020-2030	4	
11	Riverside-San Bernardino-Ontario MSA (Riversid…	2020-2030	4	
12	Bakersfield MSA (Kern County)	2020-2030	4	
13	Chico MSA (Butte County)	2020-2030	4	
14	Santa Cruz-Watsonville MSA (Santa Cruz County)	2020-2030	4	
15	Oakland-Hayward-Berkeley MD (Alameda and Contr…	2020-2030	4	

	Occupational Title	\
0	Computer and Information Systems Managers	
1	Computer Systems Analysts	
2	Computer and Information Research Scientists	
3	Computer Network Support Specialists	
4	Computer User Support Specialists	
5	Computer Network Architects	
6	Network and Computer Systems Administrators	
7	Computer Programmers	
8	Computer Occupations, All Other	
9	Computer Hardware Engineers	
10	Electronics Engineers, Except Computer	
11	Computer Science Teachers, Postsecondary	
12	Office Machine Operators, Except Computer	
13	Computer, Automated Teller, and Office Machine…	
14	Computer Numerically Controlled Tool Operators	
15	Computer Numerically Controlled Tool Programmers	

	Base Year Employment Estimate	Projected Year Employment Estimate	\
0	110.0	160.0	
1	30.0	40.0	
2	2000.0	2980.0	
3	60.0	80.0	
4	280.0	350.0	
5	50.0	60.0	
6	40.0	50.0	
7	3310.0	3600.0	
8	180.0	230.0	

9	2570.0	3320.0
10	100.0	130.0
11	160.0	200.0
12	70.0	70.0
13	50.0	70.0
14	60.0	70.0
15	210.0	290.0

	Numeric Change	Percentage Change	Exits	Transfers	Total Job Openings	\
0	50.0	45.5	20.0	80.0	150.0	
1	10.0	33.3	10.0	20.0	40.0	
2	980.0	49.0	440.0	1260.0	2680.0	
3	20.0	33.3	10.0	40.0	70.0	
4	70.0	25.0	60.0	170.0	300.0	
5	10.0	20.0	10.0	30.0	50.0	
6	10.0	25.0	10.0	20.0	40.0	
7	290.0	8.8	690.0	1580.0	2560.0	
8	50.0	27.8	50.0	100.0	200.0	
9	750.0	29.2	560.0	1380.0	2690.0	
10	30.0	30.0	30.0	50.0	110.0	
11	40.0	25.0	80.0	80.0	200.0	
12	0.0	0.0	40.0	40.0	80.0	
13	20.0	40.0	20.0	40.0	80.0	
14	10.0	16.7	20.0	50.0	80.0	
15	80.0	38.1	90.0	180.0	350.0	

	Median Hourly Wage	Median Annual Wage	Entry Level Education	\
0	65.42	136088.0	Bachelor's degree	
1	47.05	97880.0	Bachelor's degree	
2	98.08	204011.0	Master's degree	
3	30.54	63527.0	Associate's degree	
4	29.82	62014.0	Some college, no degree	
5	48.68	101258.0	Bachelor's degree	
6	48.09	100025.0	Bachelor's degree	
7	60.05	124906.0	Bachelor's degree	
8	36.75	76435.0	Bachelor's degree	
9	95.52	198681.0	Bachelor's degree	
10	49.70	103371.0	Bachelor's degree	
11	0.00	126702.0	Doctoral or professional degree	
12	18.86	39221.0	High school diploma or equivalent	
13	18.62	38727.0	Some college, no degree	
14	28.98	60289.0	High school diploma or equivalent	
15	44.34	92220.0	Postsecondary non-degree award	

	Work Experience	Job Training
0	5 years or more	None
1	None	None

```
2              None                                  None
3              None                                  None
4              None                                  None
5   5 years or more                                  None
6              None                                  None
7              None                                  None
8              None                                  None
9              None                                  None
10             None                                  None
11             None                                  None
12             None     Short-term on-the-job training
13             None     Short-term on-the-job training
14             None  Moderate-term on-the-job training
15             None  Moderate-term on-the-job training
```

[7]:
```python
# Step 4, part 1: Gather all colleges in this MSA

max_msa_colleges = pd.merge(
    all_colleges,
    max_msa_for_occupation,
    left_on="Area Name",
    right_on="Area Name",
    how="inner",
).sort_values(
    by=[
        "Standard Occupational Classification (SOC)",
        "Metropolitan Statistical Area",
        "CollegeName",
    ]
).reset_index(drop=True)

# Print relevant information:
print(f"{max_msa_colleges.shape[0]:,} rows")
max_msa_colleges[
    [
        "CollegeID",
        "CollegeName",
        "Matched Counties",
        "Area Name",
        "Metropolitan Statistical Area",
        "MSA Counties",
        "Standard Occupational Classification (SOC)",
        "Occupational Title",
    ]
].head()
```

```
53 rows
```

[7]:

	CollegeID	CollegeName	Matched Counties	Area Name	Metropolitan Statistical Area	MSA Counties	Standard Occupational Classification (SOC)	Occupational Title
0	111	Butte College	Butte County	Chico MSA (Butte County)	Chico MSA	Butte County	11-3021	Computer and Information Systems Managers
1	582	Lemoore College	Kings County	Hanford-Corcoran MSA (Kings County)	Hanford-Corcoran MSA	Kings County	15-1211	Computer Systems Analysts
2	371	Cañada College	San Francisco and San Mateo Counties	San Francisco-Redwood City-South San Francisco…	San Francisco-Redwood City-South San Francisco MD	San Francisco and San Mateo Counties	15-1221	Computer and Information Research Scientists
3	361	City College of San Francisco	San Francisco and San Mateo Counties	San Francisco-Redwood City-South San Francisco…	San Francisco-Redwood City-South San Francisco MD	San Francisco and San Mateo Counties	15-1221	Computer and Information Research Scientists
4	372	College of San Mateo	San Francisco and San Mateo Counties	San Francisco-Redwood City-South San Francisco…	San Francisco-Redwood City-South San Francisco MD	San Francisco and San Mateo Counties	15-1221	

4 Computer and Information Research Scientists

[8]:
```
# Now, gather the SOC Codes from the labor market data

max_soc_codes = (
    max_msa_colleges["Standard Occupational Classification (SOC)"].unique()
)

# These are all the SOC codes for the top MSA for the occupations which contain
# the word "Computer"
max_soc_codes
```

[8]:
```
array(['11-3021', '15-1211', '15-1221', '15-1231', '15-1232', '15-1241',
       '15-1244', '15-1251', '15-1299', '17-2061', '17-2072', '25-1021',
       '43-9071', '49-2011', '51-9161', '51-9162'], dtype=object)
```

[9]:
```
# Build a DataFrame of colleges that offer programs in the relevant TOP Codes
# for each of the top SOC Codes

matched_top_codes = pd.DataFrame()
for soc in max_soc_codes:
    df = join_tools.match_colleges_top_from_soc(soc)
    df["SocCode"] = soc
    df["SocTitle"] = occupation_by_msa.loc[
        occupation_by_msa["Standard Occupational Classification (SOC)"] == soc,
        "Occupational Title"
    ].values[0]
    matched_top_codes = pd.concat([matched_top_codes, df])

print(f"{matched_top_codes.shape[0]:,} rows")
matched_top_codes.head()
```

```
Error thrown when attempting SOC-TOP Merge with soc = '17-2072':
KeyError 'TopCode'
Error thrown when attempting SOC-TOP Merge with soc = '43-9071':
KeyError 'TopCode'
Error thrown when attempting SOC-TOP Merge with soc = '51-9161':
KeyError 'TopCode'
Error thrown when attempting SOC-TOP Merge with soc = '51-9162':
KeyError 'TopCode'
9,093 rows
```

[9]:

	CollegeID	CollegeName	ProgramAward	CreditType	\
0	021	Cuyamaca College	X	C	
1	021	Cuyamaca College	S	C	
2	021	Cuyamaca College	T	C	
3	021	Cuyamaca College	S	C	
4	021	Cuyamaca College	T	C	

```
                                                 Title TopCode  \
0                                   Computer Science  070600
1  Networking, Security and System Administration…  070810
2  Networking, Security and System Administration…  070810
3  Networking, Security and System Administration…  070810
4  Networking, Security and System Administration…  070810

                 TopCodeTitle  SocCode  \
0  Computer Science (Transfer)  11-3021
1          Computer Networking  11-3021
2          Computer Networking  11-3021
3          Computer Networking  11-3021
4          Computer Networking  11-3021

                                   SocTitle
0  Computer and Information Systems Managers
1  Computer and Information Systems Managers
2  Computer and Information Systems Managers
3  Computer and Information Systems Managers
4  Computer and Information Systems Managers
```

Question: Per MSA, which occupations are growing the fastest?

[10]:
```
top_occupation_percent_change = labor_market.
  ↪get_top_occupation_per_msa_by_percentage_change()
top_occupation_percent_change.head()
```

[10]:
```
    _id                                          _full_text          Area Type  \
0   380  '-2030':13 '-3091':16 '10450':25 '11990':26 '1…  Metropolitan Area
1   896  '-1171':11 '-2030':8 '130':19 '155710':22 '202…  Metropolitan Area
2  2088  '-1256':11 '-2030':8 '0':29 '0.00':28 '15':10 …  Metropolitan Area
3  2946  '-2030':9 '-3021':12 '100':24 '133.3':21 '2020…  Metropolitan Area
4  3547  '-2030':8 '-5092':11 '1020':16 '1240':21 '15.0…  Metropolitan Area

                                    Area Name     Period SOC Level  \
0  Anaheim-Santa Ana-Irvine MD (Orange County)  2020-2030         4
1                Bakersfield MSA (Kern County)  2020-2030         4
2                     Chico MSA (Butte County)  2020-2030         4
3             El Centro MSA (Imperial County)  2020-2030         4
4                  Fresno MSA (Fresno County)  2020-2030         4

  Standard Occupational Classification (SOC)  \
0                                    39-3091
1                                    29-1171
2                                    15-1256
3                                    25-3021
4                                    39-5092
```

```
                                  Occupational Title  \
0                Amusement and Recreation Attendants
1                                Nurse Practitioners
2  Software Developers and Software Quality Assur…
3                 Self-Enrichment Education Teachers
4                        Manicurists and Pedicurists

   Base Year Employment Estimate Projected Year Employment Estimate  \
0                           7310                               14110
1                            310                                 520
2                            250                                 410
3                             30                                  70
4                            560                                1020

   Numeric Change Percentage Change  Exits Transfers Total Job Openings  \
0            6800              93.0  10450     11990               29240
1             210              67.7     90       130                 430
2             160              64.0     80       160                 400
3              40             133.3     30        30                 100
4             460              82.1    380       400                1240

   Median Hourly Wage Median Annual Wage              Entry Level Education  \
0               15.99              33261   No formal educational credential
1               74.86             155710                    Master's degree
2                0.00                  0                  Bachelor's degree
3               21.48              44672  High school diploma or equivalent
4               15.00              31200     Postsecondary non-degree award

     Work Experience                    Job Training
0               None  Short-term on-the-job training
1               None                            None
2               None                            None
3  Less than 5 years                            None
4               None                            None
```

[11]:
```
top_occupation_numeric_change = labor_market.
↪get_top_occupation_per_msa_by_numeric_change()
top_occupation_numeric_change.head()
```

[11]:
```
    _id                                         _full_text          Area Type  \
0   351  '-2030':13 '-3023':16 '12790':24 '15.99':29 '2…  Metropolitan Area
1  1222  '-2030':8 '-7062':11 '13720':26 '18.95':27 '20…  Metropolitan Area
2  2209  '-1120':11 '-2030':8 '15.29':25 '2020':7 '2120…  Metropolitan Area
3  2991  '-1120':12 '-2030':9 '10390':25 '15.18':26 '20…  Metropolitan Area
4  3468  '-1120':11 '-2030':8 '15.21':25 '15810':23 '17…  Metropolitan Area
```

```
                                    Area Name     Period SOC Level  \
0  Anaheim-Santa Ana-Irvine MD (Orange County)  2020-2030         4
1                Bakersfield MSA (Kern County)  2020-2030         4
2                     Chico MSA (Butte County)  2020-2030         4
3             El Centro MSA (Imperial County)  2020-2030         4
4                  Fresno MSA (Fresno County)  2020-2030         4

  Standard Occupational Classification (SOC)  \
0                                    35-3023
1                                    53-7062
2                                    31-1120
3                                    31-1120
4                                    31-1120

                                Occupational Title  \
0                     Fast Food and Counter Workers
1  Laborers and Freight, Stock, and Material Move…
2                Home Health and Personal Care Aides
3                Home Health and Personal Care Aides
4                Home Health and Personal Care Aides

  Base Year Employment Estimate Projected Year Employment Estimate  \
0                         35460                              48250
1                          6120                               9580
2                          5170                               7290
3                          5770                               7910
4                         24060                              31750

  Numeric Change Percentage Change  Exits Transfers Total Job Openings  \
0          12790               36.1  40590     44250              97630
1           3460               56.5   3430      6830              13720
2           2120               41.0   3990      3530               9640
3           2140               37.1   4380      3870              10390
4           7690               32.0  17870     15810              41370

  Median Hourly Wage Median Annual Wage              Entry Level Education  \
0              15.99              33261   No formal educational credential
1              18.95              39397   No formal educational credential
2              15.29              31790  High school diploma or equivalent
3              15.18              31587  High school diploma or equivalent
4              15.21              31641  High school diploma or equivalent

  Work Experience                    Job Training
0            None  Short-term on-the-job training
1            None  Short-term on-the-job training
2            None  Short-term on-the-job training
3            None  Short-term on-the-job training
```

```
4            None  Short-term on-the-job training
```

[12]:
```
# Lots of variation per MSA

top_occupation_percent_change["Occupational Title"].value_counts()
```

[12]:
```
Occupational Title
Cooks, Restaurant
6
Ushers, Lobby Attendants, and Ticket Takers
6
Manicurists and Pedicurists
4
Amusement and Recreation Attendants
3
Baggage Porters and Bellhops
1
Passenger Vehicle Drivers, Except Bus Drivers, Transit and Intercity
1
Bartenders
1
Commercial Pilots
1
Library Technicians
1
Fitness Trainers and Aerobics Instructors
1
First-Line Supervisors of Personal Service & Entertainment and Recreation
Workers, Exc Gambling Services    1
Nurse Practitioners
1
First-Line Supervisors of Gambling Services Workers
1
Gaming Surveillance Officers and Gaming Investigators
1
Self-Enrichment Education Teachers
1
Software Developers and Software Quality Assurance Analysts and Testers
1
Food Preparation and Serving Related Workers, All Other
1
Name: count, dtype: int64
```

[13]:
```
# Clear consensus

top_occupation_numeric_change["Occupational Title"].value_counts()
```

[13]:
```
Occupational Title
Home Health and Personal Care Aides                                    18
Farmworkers and Laborers, Crop, Nursery, and Greenhouse                 4
Fast Food and Counter Workers                                           2
Software Developers and Software Quality Assurance Analysts and Testers 2
Laborers and Freight, Stock, and Material Movers, Hand                  1
Industrial Truck and Tractor Operators                                  1
Heavy and Tractor-Trailer Truck Drivers                                 1
Name: count, dtype: int64
```

Visualizations

[14]:
```
# Get the SOC Code of the largest growing occupations by numeric change

occupation = "Home Health and Personal Care Aides"
top_occupation_numeric_change[
    top_occupation_numeric_change["Occupational Title"] == occupation
].head(1)["Standard Occupational Classification (SOC)"]
```

[14]:
```
2    31-1120
Name: Standard Occupational Classification (SOC), dtype: object
```

[15]:
```
occupation_by_msa = labor_market.get_occupation_projections_by_title("Home␣
 ↪Health and Personal Care Aides")
```

[16]:
```
# soc = "31-1120" Does not exist in the SOC-CIP Crosswalk. This is beacuse it is
# a combination of two SOC codes.

socs = [
    "31-1121", # Home Health Aides
    "31-1122"  # Personal Care Aides
]

matched_top_codes = pd.DataFrame()
for soc in socs:
    df = join_tools.match_colleges_top_from_soc(soc)
    df["SocCode"] = soc
    df["SocTitle"] = occupation_by_msa.loc[
        occupation_by_msa["Standard Occupational Classification (SOC)"] ==␣
 ↪"31-1120",
        "Occupational Title"
    ].values[0]
    matched_top_codes = pd.concat([matched_top_codes, df])

print(f"{matched_top_codes.shape[0]:,} rows")
matched_top_codes.head()
```

```
26 rows
```

[16]:

	CollegeID	CollegeName	ProgramAward	CreditType	\
0	031	Imperial Valley College	M	C	
1	091	Southwestern College	NIL	N	
2	261	Santa Rosa Junior College	N	N	
3	422	Foothill College	NIL	N	
4	741	Los Angeles City College	N	N	

	Title	TopCode	TopCodeTitle	\
0	Care of the Elderly for the Health Assistant	123080	Home Health Aide	
1	Quality Home Care Provider	123080	Home Health Aide	
2	AE: Personal Care Attendant	123080	Home Health Aide	
3	Certificate of Completion in Geriatric Home Aide	123080	Home Health Aide	
4	IN-HOME SUPPORTIVE SERVICES SKILLS	123080	Home Health Aide	

	SocCode	SocTitle
0	31-1121	Home Health and Personal Care Aides
1	31-1121	Home Health and Personal Care Aides
2	31-1121	Home Health and Personal Care Aides
3	31-1121	Home Health and Personal Care Aides
4	31-1121	Home Health and Personal Care Aides

[17]:
```
# Add MSA to compare to labor market data

msa_college_data = pd.merge(
    matched_top_codes,
    all_colleges[["CollegeName", "Metropolitan Statistical Area"]],
    on="CollegeName",
    how="left"
)

print(f"{msa_college_data.shape[0]:,} rows")
msa_college_data.head()
```

```
26 rows
```

[17]:

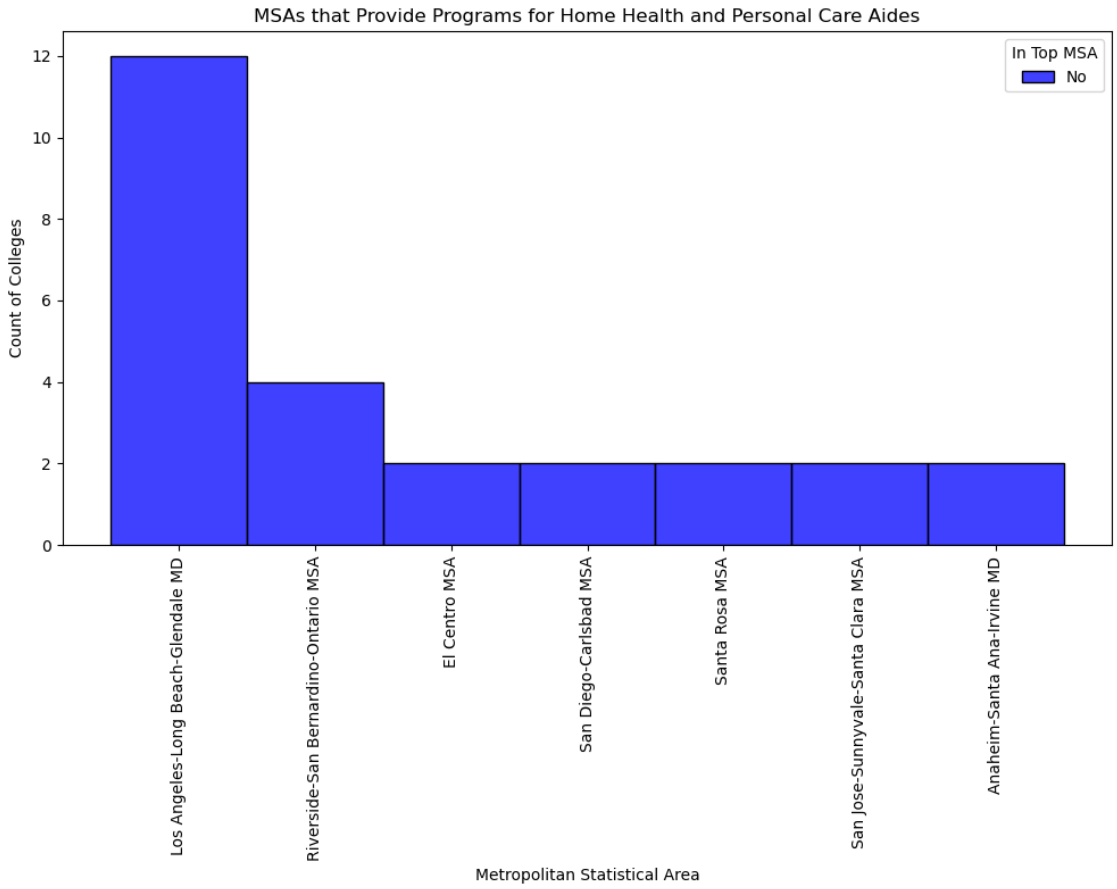
	CollegeID	CollegeName	ProgramAward	CreditType	\
0	031	Imperial Valley College	M	C	
1	091	Southwestern College	NIL	N	
2	261	Santa Rosa Junior College	N	N	
3	422	Foothill College	NIL	N	
4	741	Los Angeles City College	N	N	

	Title	TopCode	TopCodeTitle	\
0	Care of the Elderly for the Health Assistant	123080	Home Health Aide	
1	Quality Home Care Provider	123080	Home Health Aide	
2	AE: Personal Care Attendant	123080	Home Health Aide	
3	Certificate of Completion in Geriatric Home Aide	123080	Home Health Aide	
4	IN-HOME SUPPORTIVE SERVICES SKILLS	123080	Home Health Aide	

```
   SocCode                              SocTitle  \
0  31-1121  Home Health and Personal Care Aides
1  31-1121  Home Health and Personal Care Aides
2  31-1121  Home Health and Personal Care Aides
3  31-1121  Home Health and Personal Care Aides
4  31-1121  Home Health and Personal Care Aides

        Metropolitan Statistical Area
0                         El Centro MSA
1                San Diego-Carlsbad MSA
2                        Santa Rosa MSA
3  San Jose-Sunnyvale-Santa Clara MSA
4  Los Angeles-Long Beach-Glendale MD
```

[18]:
```
in_demand_areas = top_occupation_numeric_change[
    top_occupation_numeric_change["Standard Occupational Classification (SOC)"]␣
 ↪== "31-1120"
]["Area Name"].unique()

msa_college_data["In Top MSA"] = msa_college_data["Metropolitan Statistical␣
 ↪Area"].apply(
    lambda x: "Yes" if x in in_demand_areas else "No"
)
# Reorder the bars in descending order of size
msa_college_data["Metropolitan Statistical Area"] = pd.Categorical(
    msa_college_data["Metropolitan Statistical Area"],
    categories=msa_college_data["Metropolitan Statistical Area"].value_counts().
 ↪index,
    ordered=True
)

plt.gcf().set_size_inches(10, 8)
# Plot the histogram
sns.histplot(
    data=msa_college_data,
    x="Metropolitan Statistical Area",
    hue="In Top MSA",
    multiple="stack",
    palette={"Yes": "green", "No": "blue"},
    discrete=True
)
plt.xticks(rotation=90)
plt.title("MSAs that Provide Programs for Home Health and Personal Care Aides")
plt.xlabel("Metropolitan Statistical Area")
plt.ylabel("Count of Colleges")
plt.tight_layout()
```

```
plt.savefig("figures/msa_in_demand.png")
```



Now we consider popular majors/programs vs labor market trends. According to DataMart, *psychology* is the most popular program, other than the *biological and physical sciences (and mathematics)* joint programs.

[19]:
```
# Sorted by percent change.

psych_jobs = labor_market.get_occupation_projections_by_title("Psych")
psych_jobs.head()
```

[19]:
```
    _id                                          _full_text  \
0   3855  '-2030':10 '-3031':13 '0':27 '0.00':26 '100':2…
1  14771  '-2030':9 '-3031':12 '0':26 '0.00':25 '120':23…
2   8769  '-2030':8 '-3031':11 '0':25 '0.00':24 '10':19 …
3  15784  '-2030':10 '-3031':13 '0':27 '0.00':26 '100':2…
4  11029  '-2030':12 '-3031':15 '0':29 '0.00':28 '15.3':…

           Area Type                                  Area Name     Period  \
```

```
0  Metropolitan Area        Hanford-Corcoran MSA (Kings County)  2020-2030
1  Metropolitan Area             Santa Rosa MSA (Sonoma County)  2020-2030
2  Metropolitan Area                Redding MSA (Shasta County)  2020-2030
3  Metropolitan Area      Vallejo-Fairfield MSA (Solano County)  2020-2030
4  Metropolitan Area  San Diego-Carlsbad MSA (San Diego County)  2020-2030

  SOC Level Standard Occupational Classification (SOC)  \
0         4                                    19-3031
1         4                                    19-3031
2         4                                    19-3031
3         4                                    19-3031
4         4                                    19-3031

                              Occupational Title  \
0  Clinical, Counseling, and School Psychologists
1  Clinical, Counseling, and School Psychologists
2  Clinical, Counseling, and School Psychologists
3  Clinical, Counseling, and School Psychologists
4  Clinical, Counseling, and School Psychologists

   Base Year Employment Estimate  Projected Year Employment Estimate  \
0                          110.0                                130.0
1                          290.0                                340.0
2                           60.0                                 70.0
3                          240.0                                280.0
4                         1770.0                               2040.0

   Numeric Change  Percentage Change  Exits  Transfers  Total Job Openings  \
0            20.0               18.2   30.0       50.0               100.0
1            50.0               17.2   80.0      120.0               250.0
2            10.0               16.7   20.0       30.0                60.0
3            40.0               16.7   70.0      100.0               210.0
4           270.0               15.3  500.0      750.0              1520.0

   Median Hourly Wage  Median Annual Wage            Entry Level Education  \
0                 0.0                 0.0  Doctoral or professional degree
1                 0.0                 0.0  Doctoral or professional degree
2                 0.0                 0.0  Doctoral or professional degree
3                 0.0                 0.0  Doctoral or professional degree
4                 0.0                 0.0  Doctoral or professional degree

  Work Experience          Job Training
0            None  Internship/residency
1            None  Internship/residency
2            None  Internship/residency
3            None  Internship/residency
4            None  Internship/residency
```

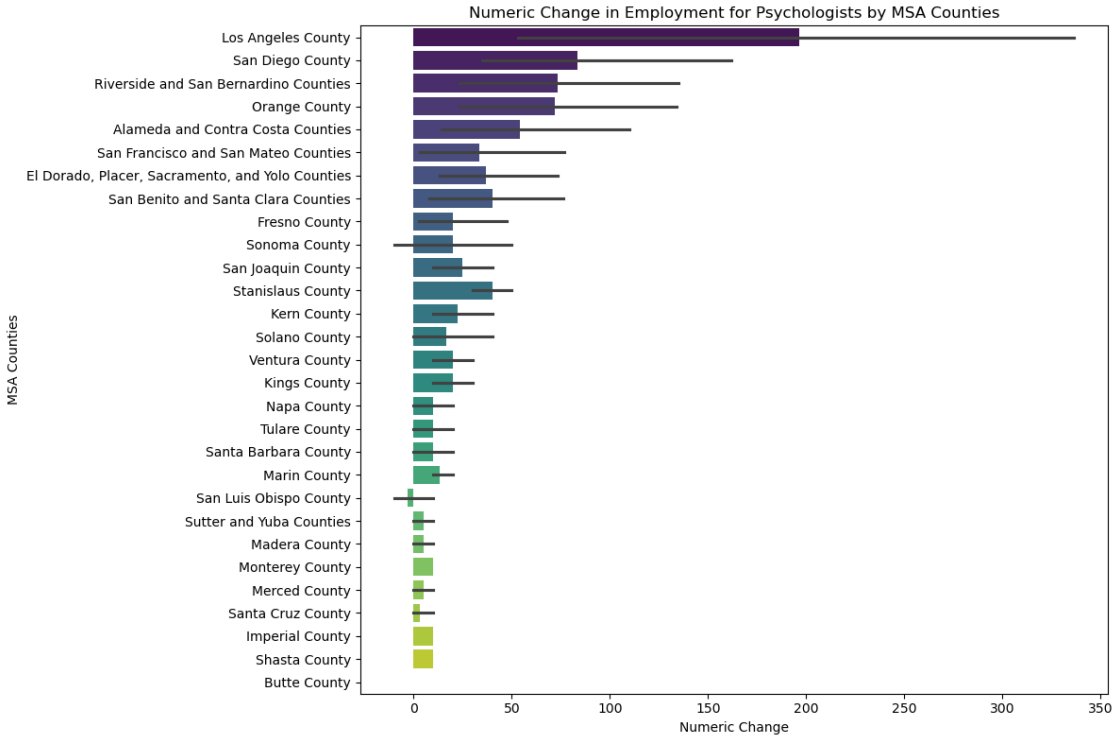
[20]:
```
# Sorted by numeric change.

psych_jobs.sort_values(by="Numeric Change", ascending=False).head()
```

[20]:
```
      _id                                          _full_text  \
19   4287  '-2030':15 '-3031':18 '0':32 '0.00':31 '1310':…
78   4479  '-2030':15 '-2053':18 '1240':26 '13.7':24 '202…
4   11029  '-2030':12 '-3031':15 '0':29 '0.00':28 '15.3':…
68   4464  '-1223':18 '-2030':15 '0':28 '0.00':27 '14.4':…
18   9225  '-2030':16 '-3031':19 '0':33 '0.00':32 '10.0':…

            Area Type                                          Area Name  \
19  Metropolitan Area  Los Angeles-Long Beach-Glendale MD (Los Angele…
78  Metropolitan Area  Los Angeles-Long Beach-Glendale MD (Los Angele…
4   Metropolitan Area          San Diego-Carlsbad MSA (San Diego County)
68  Metropolitan Area  Los Angeles-Long Beach-Glendale MD (Los Angele…
18  Metropolitan Area  Riverside-San Bernardino-Ontario MSA (Riversid…

       Period SOC Level Standard Occupational Classification (SOC)  \
19  2020-2030         4                                     19-3031
78  2020-2030         4                                     29-2053
4   2020-2030         4                                     19-3031
68  2020-2030         4                                     29-1223
18  2020-2030         4                                     19-3031

                              Occupational Title  \
19  Clinical, Counseling, and School Psychologists
78                        Psychiatric Technicians
4   Clinical, Counseling, and School Psychologists
68                                  Psychiatrists
18  Clinical, Counseling, and School Psychologists

    Base Year Employment Estimate  Projected Year Employment Estimate  \
19                         4790.0                               5260.0
78                         2710.0                               3080.0
4                          1770.0                               2040.0
68                         1600.0                               1830.0
18                         2000.0                               2200.0

    Numeric Change  Percentage Change   Exits  Transfers  Total Job Openings  \
19           470.0                9.8  1310.0     1960.0              3740.0
78           370.0               13.7   790.0     1240.0              2400.0
4            270.0               15.3   500.0      750.0              1520.0
68           230.0               14.4   280.0      190.0               700.0
18           200.0               10.0   550.0      820.0              1570.0

    Median Hourly Wage  Median Annual Wage            Entry Level Education  \
```

```
19                0.00                0.0  Doctoral or professional degree
78               23.02            47876.0   Postsecondary non-degree award
4                 0.00                0.0  Doctoral or professional degree
68                0.00                0.0  Doctoral or professional degree
18                0.00                0.0  Doctoral or professional degree

      Work Experience                    Job Training
19               None            Internship/residency
78  Less than 5 years  Short-term on-the-job training
4                None            Internship/residency
68               None            Internship/residency
18               None            Internship/residency
```

[21]:
```
sorted_psych_jobs = psych_jobs.sort_values(by="Numeric Change", ascending=False)


sorted_psych_jobs = sorted_psych_jobs.merge(
    msa[["Area Name", "MSA Counties"]], on="Area Name", how="left"
)

plt.figure(figsize=(12, 8))
sns.barplot(
    data=sorted_psych_jobs,
    x="Numeric Change",
    y="MSA Counties",
    palette="viridis",
)
plt.title("Numeric Change in Employment for Psychologists by MSA Counties")
plt.xlabel("Numeric Change")
plt.ylabel("MSA Counties")
plt.tight_layout()
plt.savefig("figures/psychologist_employment.png")
plt.show()
```

```
C:\Users\codyc\AppData\Local\Temp\ipykernel_18916\248909234.py:9: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(
```


Numeric Change in Employment for Psychologists by MSA Counties
Los Angeles County
San Diego County
Riverside and San Bernardino Counties
Orange County
Alameda and Contra Costa Counties
San Francisco and San Mateo Counties
El Dorado, Placer, Sacramento, and Yolo Counties
San Benito and Santa Clara Counties
Fresno County
Sonoma County
San Joaquin County
Stanislaus County
Kern County
Solano County
Ventura County
Kings County
Napa County
Tulare County
Santa Barbara County
Marin County
San Luis Obispo County
Sutter and Yuba Counties
Madera County
Monterey County
Merced County
Santa Cruz County
Imperial County
Shasta County
Butte County
MSA Counties
0
50
100
150
200
250
300
350
Numeric Change