In [2]:
```
#DATA MANIPULATION
import pandas as pd
import geopandas as gpd

#DATA VISUALIZATION
import matplotlib.pyplot as plt

# plots to get appear and store within notebook
%matplotlib inline
```

In [3]:
```
# reading csv REI location data
rei_loc = pd.read_csv(r'G:\BCG CASE STUDY\rei_case_study_data\rei_data\rei_locations.csv')
```

In [4]:
```
rei_loc.head()
```

Out[4]:

	REI_STORE_ID	list_name	store_name	address	city	state	zip_code	phone_number	latitude	longitude	...	source_url	spe
0	REI_1	REI	REI Seattle flagship store	222 Yale Ave N	Seattle	WA	98109	(206) 223-1944	47.620207	-122.329518	...	www.rei.com	
1	REI_2	REI	REI Berkeley store	1338 San Pablo Ave	Berkeley	CA	94702	(510) 527-4140	37.878994	-122.296276	...	www.rei.com	
2	REI_3	REI	REI Portland store	1405 NW Johnson St	Portland	OR	97209	(503) 221-1938	45.528917	-122.685760	...	www.rei.com	
3	REI_4	REI	REI Manhattan Beach store	1800 Rosecrans Ave	Manhattan Beach	CA	90266	(310) 727-0728	33.901706	-118.380407	...	www.rei.com	
4	REI_5	REI	REI Bloomington flagship store	750 W American Blvd	Bloomington	MN	55420	(952) 884-4315	44.860997	-93.291179	...	www.rei.com	

5 rows × 52 columns

In [5]:
```
# reading csv REI performance data
rei_rev = pd.read_csv(r'G:\BCG CASE STUDY\rei_case_study_data\rei_data\rei_performance_metricss.csv')
```

In [6]:
```
rei_rev.head()
```

Out[6]:

	REI_STORE_ID	revenue_year_2016	average_customer_spend / year_2016	revenue_year_2021	average_customer_spend / year_2021
0	REI_1	29255444.23	33.624899	34228335.75	37.301331
1	REI_2	24402937.84	21.140366	27827459.82	23.373382
2	REI_3	28502739.52	36.791353	33333291.66	40.426482
3	REI_4	50484881.23	20.022504	56572934.88	22.005272
4	REI_5	23237085.49	25.887326	25756405.74	27.567095

In [7]:
```
# Merge the both DATAFRAMES based on REI STORE ID

merge_rei = pd.merge(rei_loc,rei_rev,on ="REI_STORE_ID" )
```

In [8]:
```
merge_rei.head()
```

Out[8]:

	REI_STORE_ID	list_name	store_name	address	city	state	zip_code	phone_number	latitude	longitude	...	storefront	tena
0	REI_1	REI	REI Seattle flagship store	222 Yale Ave N	Seattle	WA	98109	(206) 223-1944	47.620207	-122.329518	...	REI - REI Seattle flagship store	
1	REI_2	REI	REI Berkeley store	1338 San Pablo Ave	Berkeley	CA	94702	(510) 527-4140	37.878994	-122.296276	...	REI - REI Berkeley store	
2	REI_3	REI	REI Portland store	1405 NW Johnson St	Portland	OR	97209	(503) 221-1938	45.528917	-122.685760	...	REI - REI Portland store	
3	REI_4	REI	REI Manhattan Beach store	1800 Rosecrans Ave	Manhattan Beach	CA	90266	(310) 727-0728	33.901706	-118.380407	...	REI - REI Manhattan Beach store	
4	REI_5	REI	REI Bloomington flagship store	750 W American Blvd	Bloomington	MN	55420	(952) 884-4315	44.860997	-93.291179	...	REI - REI Bloomington flagship store	

5 rows × 56 columns

In [9]:
```
# reading shapefile
demographic = gpd.read_file(r"G:\BCG CASE STUDY\rei_case_study_data\GAMMA_demographics\zip_demographics.shp")
```

In [10]:
```
demographic.head()
```

Out[10]:

	Name	Key	XCYA04V001	XCYA04V002	XCYA04V003	XCYA04V004	XCYA04T18P	XCYA04V005	XCYA04V006	XCYA04V007	...	CYEC2
0	00001 N Dillingham Census Area	00001	11.13	8.99	8.75	5.24	65.88	3.72	1.78	1.97	...	
1	00002 Yukon Flats Nat Wildlife	00002	9.97	7.57	7.32	4.32	70.82	3.11	1.51	1.53	...	
2	00003 Alaska Peninsula NWR	00003	7.05	6.27	6.11	4.55	76.02	2.66	1.25	1.25	...	
3	00004 W Kenai Peninsula Boroug	00004	3.15	3.41	2.89	2.10	88.45	0.52	1.31	1.05	...	
4	00005 N Lake and Peninsula Bor	00005	10.56	8.80	8.49	4.45	67.70	2.80	1.86	1.76	...	

5 rows × 63 columns

In [11]:
```
# Rename the column
demographic.rename({"Key":"zip_code"},axis=1, inplace=True)
```

In [12]:
```
demographic.head()
```

Out[12]:

	Name	zip_code	XCYA04V001	XCYA04V002	XCYA04V003	XCYA04V004	XCYA04T18P	XCYA04V005	XCYA04V006	XCYA04V007	...	CYI
0	00001 N Dillingham Census Area	00001	11.13	8.99	8.75	5.24	65.88	3.72	1.78	1.97	...	
1	00002 Yukon Flats Nat Wildlife	00002	9.97	7.57	7.32	4.32	70.82	3.11	1.51	1.53	...	
2	00003 Alaska Peninsula NWR	00003	7.05	6.27	6.11	4.55	76.02	2.66	1.25	1.25	...	
3	00004 W Kenai Peninsula Boroug	00004	3.15	3.41	2.89	2.10	88.45	0.52	1.31	1.05	...	
4	00005 N Lake and Peninsula Bor	00005	10.56	8.80	8.49	4.45	67.70	2.80	1.86	1.76	...	

5 rows × 63 columns

In [13]:
```
# Merging the geodataframe with the dataframe
merge_data = pd.merge(merge_rei,demographic,on="zip_code")
```

In [14]:
```
merge_data.head()
```

Out[14]:

	REI_STORE_ID	list_name	store_name	address	city	state	zip_code	phone_number	latitude	longitude	...	CYEC20V003	CYI
0	REI_1	REI	REI Seattle flagship store	222 Yale Ave N	Seattle	WA	98109	(206) 223-1944	47.620207	-122.329518	...	90305	
1	REI_2	REI	REI Berkeley store	1338 San Pablo Ave	Berkeley	CA	94702	(510) 527-4140	37.878994	-122.296276	...	101242	
2	REI_3	REI	REI Portland store	1405 NW Johnson St	Portland	OR	97209	(503) 221-1938	45.528917	-122.685760	...	68244	
3	REI_4	REI	REI Manhattan Beach store	1800 Rosecrans Ave	Manhattan Beach	CA	90266	(310) 727-0728	33.901706	-118.380407	...	186189	
4	REI_5	REI	REI Bloomington flagship store	750 W American Blvd	Bloomington	MN	55420	(952) 884-4315	44.860997	-93.291179	...	62649	

5 rows × 118 columns

In [15]:
```
# Save the data to csv file for further analysis
merge_data.to_csv("G:\BCG CASE STUDY\merge_data.csv")
```

In [77]:
```
# Comparing Revenue with Population density data)
PD1 = merge_data[["CYPOPDENS","revenue_year_2016"]]
```

In [78]:
```
# Sorting data based on Revenue from high to low value
PD2 = PD1.sort_values('revenue_year_2016',ascending=False)
```

In [79]:
```
# Inorder to compare the first 30 datasets
PD3 = PD2[:30]
PD3.head()
```

Out[79]:

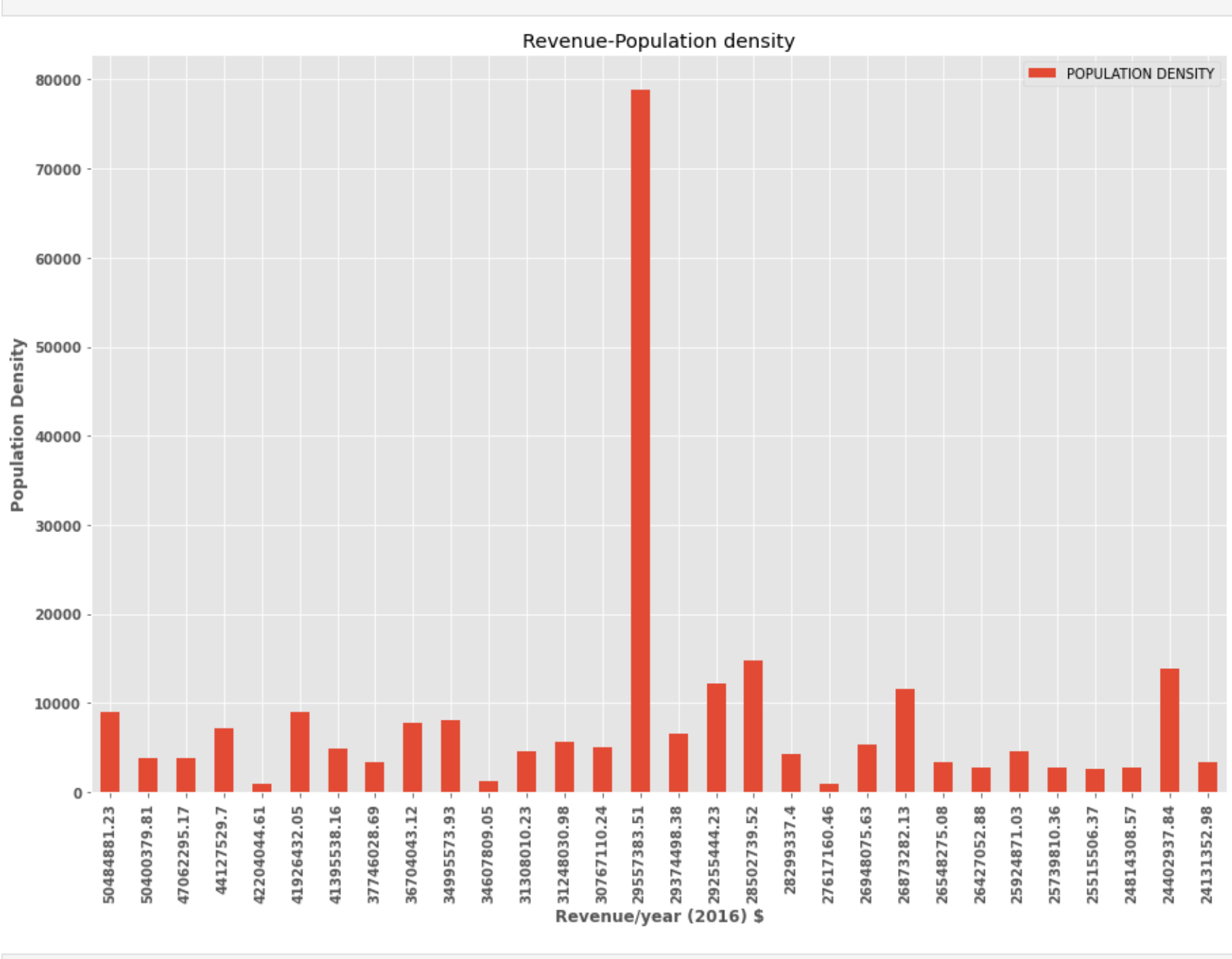
	CYPOPDENS	revenue_year_2016
3	9030.0	50484881.23
63	3783.0	50400379.81
77	3858.0	47062295.17
30	7154.0	44127529.70
9	858.8	42204044.61

In [80]:
```
# Data visualization

# Preparing Revenue- Population Bar plot
PD3.plot.bar(x="revenue_year_2016",y="CYPOPDENS",label="POPULATION DENSITY",fontsize=20)
import matplotlib
matplotlib.style.use('ggplot')

plt.rcParams['figure.figsize'] =(15,10)
plt.xlabel("Revenue/year (2016) $",fontsize=12,fontweight='bold')
plt.ylabel("Population Density",fontsize=12,fontweight='bold')
plt.title("Revenue-Population density")
plt.xticks(fontsize=10,fontweight='bold')
plt.yticks(fontsize=10,fontweight='bold')

plt.savefig("Revenue-Population density statistics chart.png",bbox_inches="tight")
```



In [81]:
```
# Comparing Revenue with Total Population data)
TP1 = merge_data[["CYA01V001","revenue_year_2016"]]
```

In [82]:
```
# Sorting data based on Revenue from high to low value
TP2 = TP1.sort_values('revenue_year_2016',ascending=False)
```

In [83]:
```
# Inorder to compare the first 30 datasets
TP3 = TP2[:30]
TP3.head()
```

Out[83]:

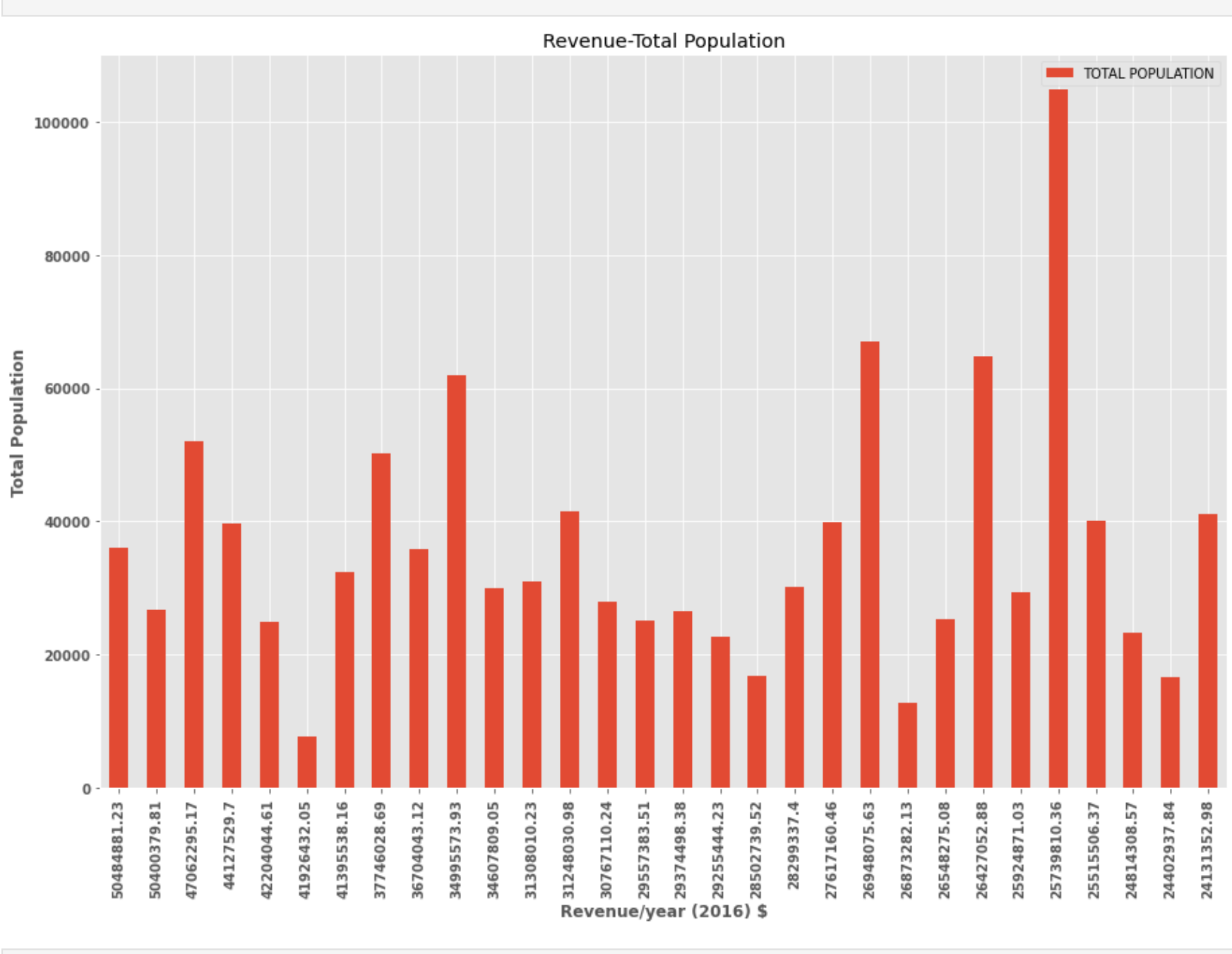
	CYA01V001	revenue_year_2016
3	35928	50484881.23
63	26665	50400379.81
77	52005	47062295.17
30	39596	44127529.70
9	24821	42204044.61

In [84]:
```
# Data visualization

# Preparing Revenue- Population Bar plot
TP3.plot.bar(x="revenue_year_2016",y="CYA01V001",label="TOTAL POPULATION",fontsize=20)
import matplotlib
matplotlib.style.use('ggplot')

plt.rcParams['figure.figsize'] =(15,10)
plt.xlabel("Revenue/year (2016) $",fontsize=12,fontweight='bold')
plt.ylabel("Total Population",fontsize=12,fontweight='bold')
plt.title("Revenue-Total Population")
plt.xticks(fontsize=10,fontweight='bold')
plt.yticks(fontsize=10,fontweight='bold')

plt.savefig("Revenue-Total Population statistics chart.png",bbox_inches="tight")
```



In [86]:
```
# Comparing Revenue with Consumer Expenditure data)
ex1 = merge_data[["FATOTALEXP","revenue_year_2021"]]
```

In [87]:
```
# Sorting data based on Revenue from high to low value
ex2 = ex1.sort_values('revenue_year_2021',ascending=False)
```

In [88]:
```
# Inorder to compare the first 30 datasets
ex3 = ex2[:30]
ex3.head()
```

Out[88]:

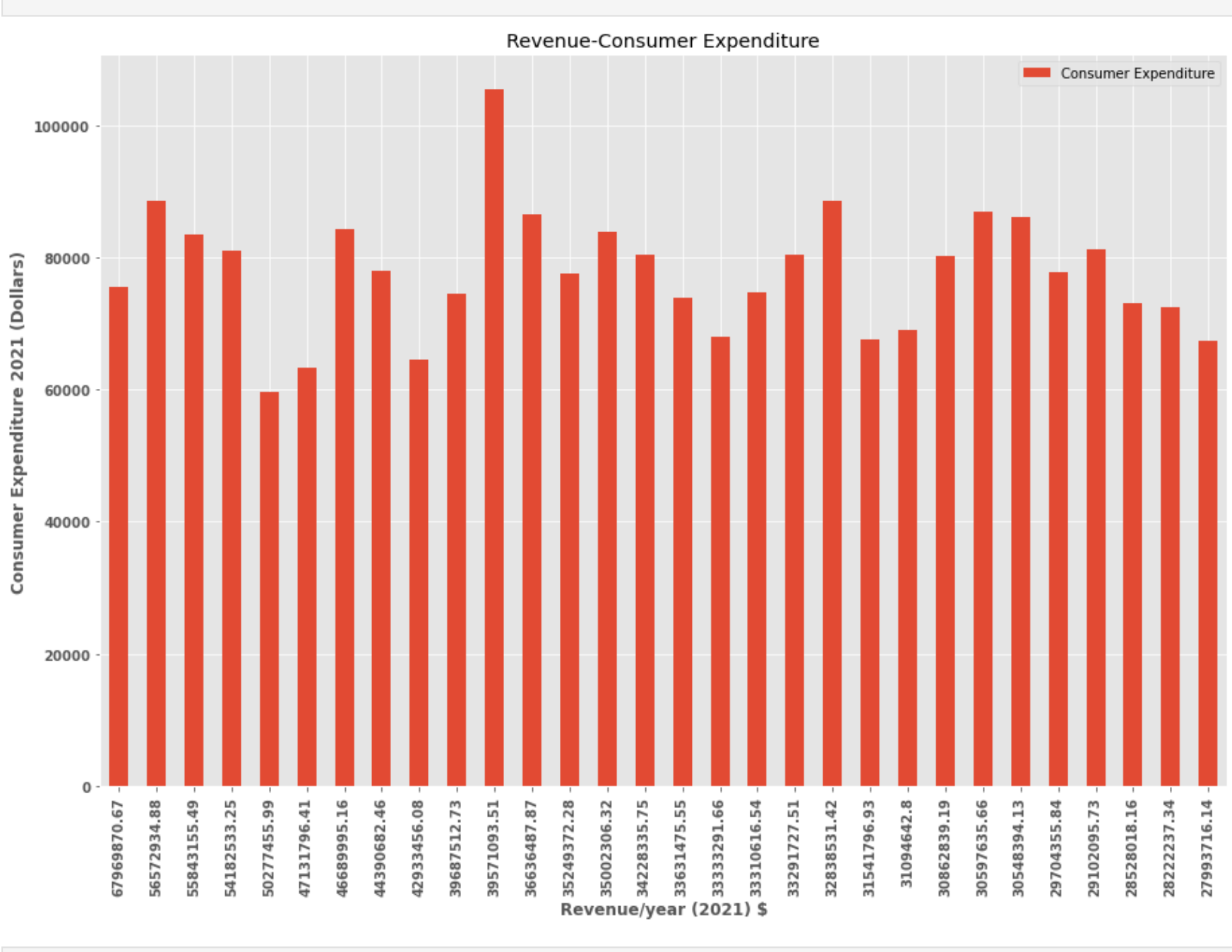
	FATOTALEXP	revenue_year_2021
63	75543.22	67969870.67
3	88594.26	56572934.88
9	83533.73	55843155.49
77	81108.20	54182533.25
30	59659.13	50277455.99

In [89]:
```
# Data visualization

# Preparing Revenue- Consumer Expenditure Bar plot
ex3.plot.bar(x="revenue_year_2021",y="FATOTALEXP",label="Consumer Expenditure",fontsize=20)
import matplotlib
matplotlib.style.use('ggplot')

plt.rcParams['figure.figsize'] =(15,10)
plt.xlabel("Revenue/year (2021) $",fontsize=12,fontweight='bold')
plt.ylabel("Consumer Expenditure 2021 (Dollars)",fontsize=12,fontweight='bold')
plt.title("Revenue-Consumer Expenditure")
plt.xticks(fontsize=10,fontweight='bold')
plt.yticks(fontsize=10,fontweight='bold')

plt.savefig("Revenue-Consumer Expenditure statistics chart.png",bbox_inches="tight")
```



In [91]:
```
# Comparing Revenue with Median Household Income data)
MH1 = merge_data[["CYEC17V001","revenue_year_2016"]]
```

In [92]:
```
# Sorting data based on Revenue from high to low value
MH2 = MH1.sort_values('revenue_year_2016',ascending=False)
```

In [93]:
```
# Inorder to compare the first 30 datasets
MH3 = MH2[:30]
MH3.head()
```

Out[93]:

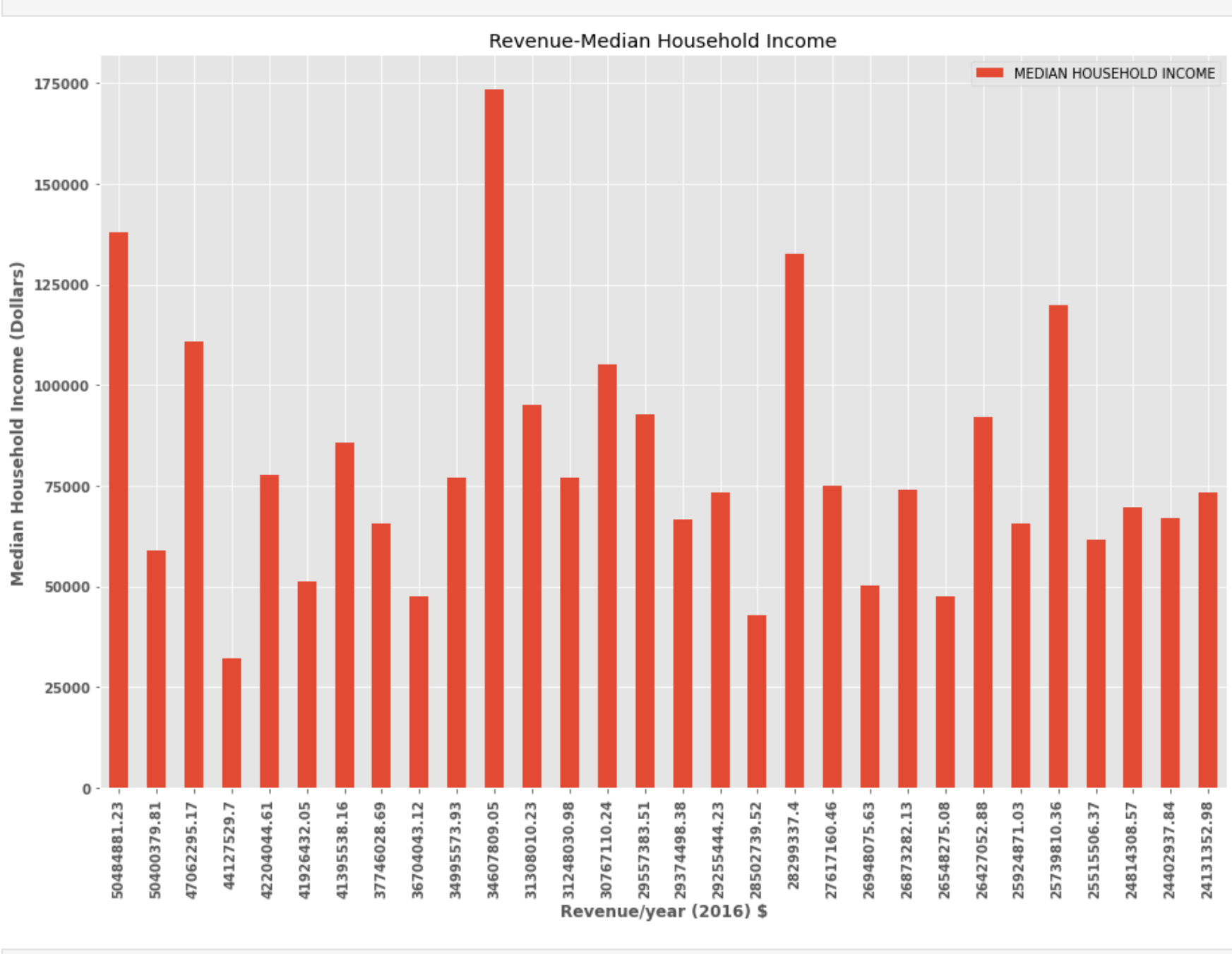
	CYEC17V001	revenue_year_2016
3	137789	50484881.23
63	58889	50400379.81
77	110797	47062295.17
30	32171	44127529.70
9	77562	42204044.61

In [94]:
```
# Data visualization

# Preparing Revenue- Median household Income Bar plot
MH3.plot.bar(x="revenue_year_2016",y="CYEC17V001",label="MEDIAN HOUSEHOLD INCOME",fontsize=20)
import matplotlib
matplotlib.style.use('ggplot')

plt.rcParams['figure.figsize'] =(15,10)
plt.xlabel("Revenue/year (2016) $",fontsize=12,fontweight='bold')
plt.ylabel("Median Household Income (Dollars)",fontsize=12,fontweight='bold')
plt.title("Revenue-Median Household Income")
plt.xticks(fontsize=10,fontweight='bold')
plt.yticks(fontsize=10,fontweight='bold')

plt.savefig("Revenue-Median Household Income statistics chart.png",bbox_inches="tight")
```



In [95]:
```
# Comparing Revenue with Bachelors Degree data)
edu1 = merge_data[["CYD02VV13","revenue_year_2016"]]
```

In [96]:
```
# Sorting data based on Revenue from high to low value
edu2 = edu1.sort_values('revenue_year_2016',ascending=False)
```

In [97]:
```
# Inorder to compare the first 30 datasets
edu3 = edu2[:30]
edu3.head()
```

Out[97]:

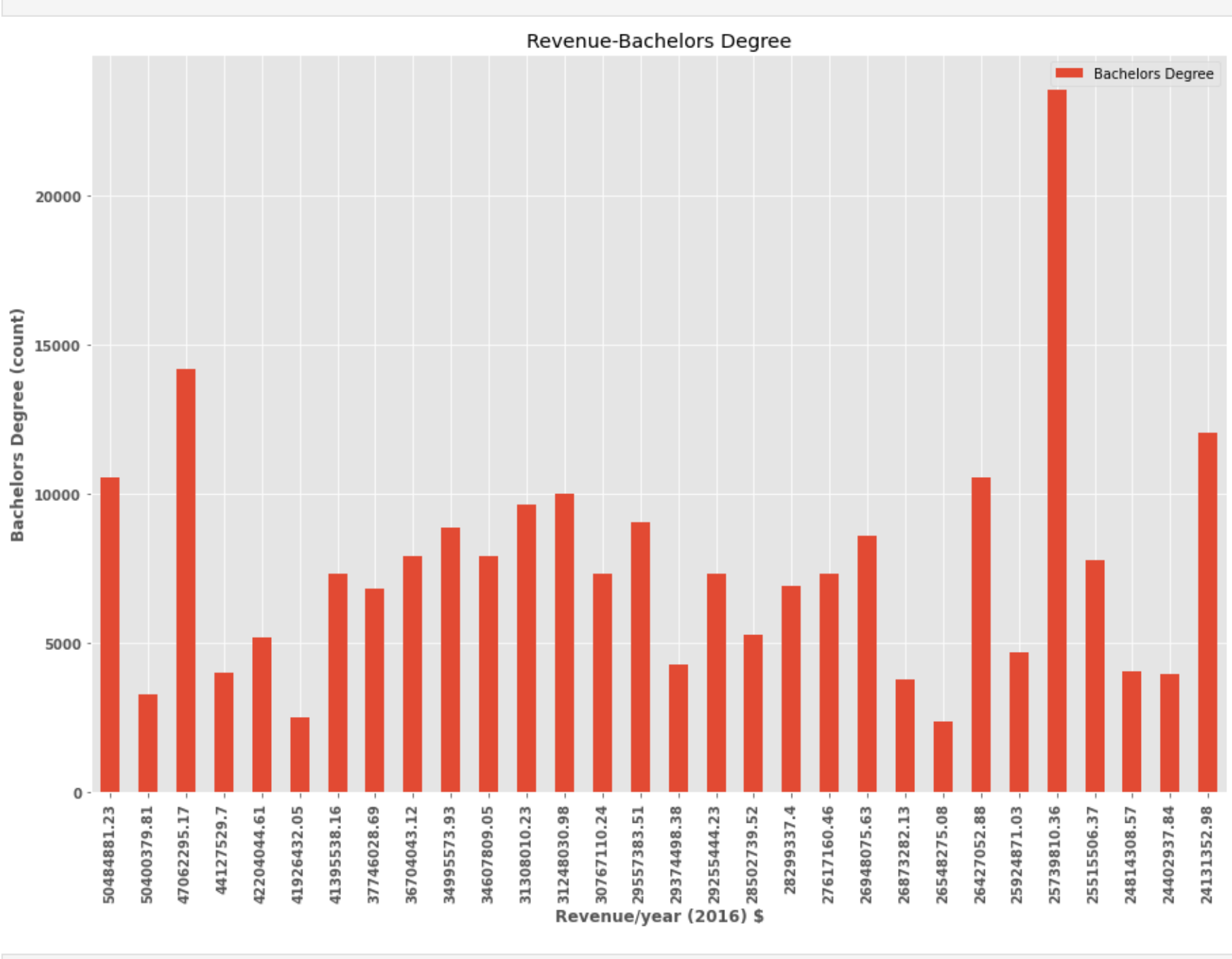
	CYD02VV13	revenue_year_2016
3	10536	50484881.23
63	3271	50400379.81
77	14195	47062295.17
30	3981	44127529.70
9	5188	42204044.61

In [98]:
```
# Data visualization

# Preparing Revenue- Bachelors degree Bar plot
edu3.plot.bar(x="revenue_year_2016",y="CYD02VV13",label="Bachelors Degree",fontsize=20)
import matplotlib
matplotlib.style.use('ggplot')

plt.rcParams['figure.figsize'] =(15,10)
plt.xlabel("Revenue/year (2016) $",fontsize=12,fontweight='bold')
plt.ylabel("Bachelors Degree (count)",fontsize=12,fontweight='bold')
plt.title("Revenue-Bachelors Degree")
plt.xticks(fontsize=10,fontweight='bold')
plt.yticks(fontsize=10,fontweight='bold')

plt.savefig("Revenue-Education statistics chart.png",bbox_inches="tight")
```



In [99]:
```
# Comparing Revenue with youngsters data)
age1 = merge_data[["XCYA04V009","revenue_year_2016"]]
```

In [100…
```
# Sorting data based on Revenue from high to low value
age2 = age1.sort_values('revenue_year_2016',ascending=False)
```

In [101…
```
# Inorder to compare the first 30 datasets
age3 = age2[:30]
age3.head()
```

Out[101…

	XCYA04V009	revenue_year_2016
3	5.32	50484881.23
63	8.95	50400379.81
77	5.65	47062295.17
30	10.16	44127529.70
9	7.54	42204044.61

In [102…
```
# Data visualization

# Preparing Revenue- youngsters Bar plot
age3.plot.bar(x="revenue_year_2016",y="XCYA04V009",label="Youngsters",fontsize=20)
import matplotlib
matplotlib.style.use('ggplot')

plt.rcParams['figure.figsize'] =(15,10)
plt.xlabel("Revenue/year (2016) $",fontsize=12,fontweight='bold')
plt.ylabel("Youngsters %",fontsize=12,fontweight='bold')
plt.title("Revenue-Youngsters")
plt.xticks(fontsize=10,fontweight='bold')
plt.yticks(fontsize=10,fontweight='bold')

plt.savefig("Revenue-Youngsters statistics chart.png",bbox_inches="tight")
```