

```
In [35]: import polars as pl
import matplotlib.pyplot as plt

# Load job applicant csv file into dataframe
job_applicants_df = pl.read_csv("Job_Applicants_by_Gender_and_Ethnicity.csv")
assert job_applicants_df is not None
assert job_applicants_df.shape == (187, 14)
print(job_applicants_df.head)

pl.Config.set_tbl_cols(100)
```

<bound method DataFrame.head of shape: (187, 14)>

Fis	Job	Job	App	Fem	Mal	Unk	Black	Hispan	Asian	Cauca			
Ameri	Filip	Unkno		ale	e	now	---	ic	---	sian			
cal	Num	Des	s										
can	ino	wn_Et											
Yea	ber	cri	Rec	---	---	n_G	i64	---	i64	---			
India	---	hnici											
r	---	pti	eiv	i64	i64	end		i64			i64		
n/	i64	ty											
---	str	on	ed			er							
Alask		---											
str		---	---			---							
an		i64											
Nativ		str	i64			i64							

i64													
201	920	311	54	20	31	3	25	18	1	6			
0	0	4											
3-2	6	DIR											
014	OP	ECT											
	201	OR											
	4/0	920											
	4/1	6											
	8												

3	201	122	ACC	648	488	152	8	151	204	123	62
		79	26								
	3-2	3 P	OUN								
0	014	201	TIN								
		3/0	G								
		8/0	CLE								
0		9	RK								
			122								
			3								
0	201	726	AIR	51	13	37	1	8	12	9	20
		0	2								
	3-2	0	POR								
0	014	0P	T								
		201	MAN								
		4/0	AGE								
0		2/1	R								
		4	726								
			0								
0	201	322	AIR	48	9	38	1	21	14	3	7
		1	2								
	3-2	7 P	POR								
0	014	201	T								
		3/1	POL								
		1/1	ICE								
0		5	LIE								
			UTE								
			NAN								
0			T								

			201									
			3									
1	201	240	AQU	40	15	24	1	3	7	7	19	
	1	1	2									
	3-2	0 0	ARI									
	014	201	ST									
		4/0	240									
		5/0	0									
		2										
...	
...	201	784	WAS	16	6	9	1	3	0	5	7	
0		0	1									
	4-2	0 P	TEW									
	015	201	ATE									
		4/7	R									
		/25	TRE									
		-AR	ATM									
		CHI	ENT									
		VE	LAB									
			ORA									
			TOR									
			...									
7	201	412	WAS	125	9	113	3	29	38	10	32	
		4	5									
	4-2	3 0	TEW									
	015	201	ATE									
		4/0	R									
		7/0	TRE									

		4-A	ATM									
		RCH	ENT									
		IVE	OPE									
			RAT									
			OR									
			...									
0	201	785	WAT	179	89	82	8	13	37	64	25	
		18	22									
	4-2	7 0	ER									
	015	201	MIC									
		4/7	ROB									
		/18	IOL									
		-AR	OGI									
		CHI	ST									
		VE	785									
			7 -									
			AR...									
1	201	391	WAT	96	2	92	2	8	48	6	23	
		7	3									
	4-2	2/P	ER									
	015	/20	UTI									
		14/	LIT									
		07/	Y									
		25-	WOR									
		ARC	KER									
		HIV	391									
		E	2 -									

```
Out[35]: polars.config.Config
```

Page 5 of 8

shape: (9, 9)

statistic	Apps Unknown_ Received Ethnicity	Black	Hispanic	Asian	Caucasia	American
count	187.0	187.0	187.0	187.0	187.0	187.0
187.0	187.0					
null_count	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0					
mean	499.7219	144.5294	182.1657	37.60427	84.47058	3.374332
19.63101	27.94652					
6	25	12	75	8	8	
std	4					
61.75382	2252.044	948.2905	824.1933	98.94178	320.9112	14.93818
	123.8977					
7	225	8	75	6	91	7
min	52					
0.0	5.0	0.0	0.0	0.0	0.0	0.0
25%	0.0					
1.0	38.0	5.0	10.0	3.0	9.0	0.0
50%	2.0					
5.0	100.0	14.0	24.0	9.0	26.0	1.0
75%	6.0					
13.0	263.0	42.0	97.0	29.0	64.0	2.0
max	20.0					
740.0	28230.0	12618.0	10214.0	1094.0	3843.0	153.0
	1475.0					

Out[36]: <function __main__.stats_overview(job_applicants_df)>

```
In [37]: # Generate a table showing the total number of applicants by ethnicity
def total_and_eth_value(job_applicants_df):
    total_and_eth = job_applicants_df.select(
        [
```

```

        pl.sum("Apps Received").alias("Apps Received"),
        pl.sum("Black").alias("Black"),
        pl.sum("Hispanic").alias("Hispanic"),
        pl.sum("Asian").alias("Asian"),
        pl.sum("Caucasian").alias("Caucasian"),
        pl.sum("American Indian/ Alaskan Native").alias(
            "American Indian/ Alaskan Native"
        ),
        pl.sum("Filipino").alias("Filipino"),
        pl.sum("Unknown_Ethnicity").alias("Unknown_Ethnicity"),
    ]
)

# Add a row name for the total row
total_and_eth = total_and_eth.with_columns(pl.lit("total").alias("statistic"))
total_by_value = total_and_eth.select(
    ["statistic"] + [col for col in total_and_eth.columns if col != "statistic"]
)
return total_by_value

total_and_eth_value(job_applicants_df)

```

Out[37]: shape: (1, 9)

statistic	Apps Received	Black	Hispanic	Asian	Caucasian	American Indian/ Alaskan Native	Filipino	Unknown_
str	i64	i64	i64	i64	i64	i64	i64	
"total"	93448	27027	34065	7032	15796	631	3671	

In [38]: *# calculate total number of applicants by ethnicity, for plotting*

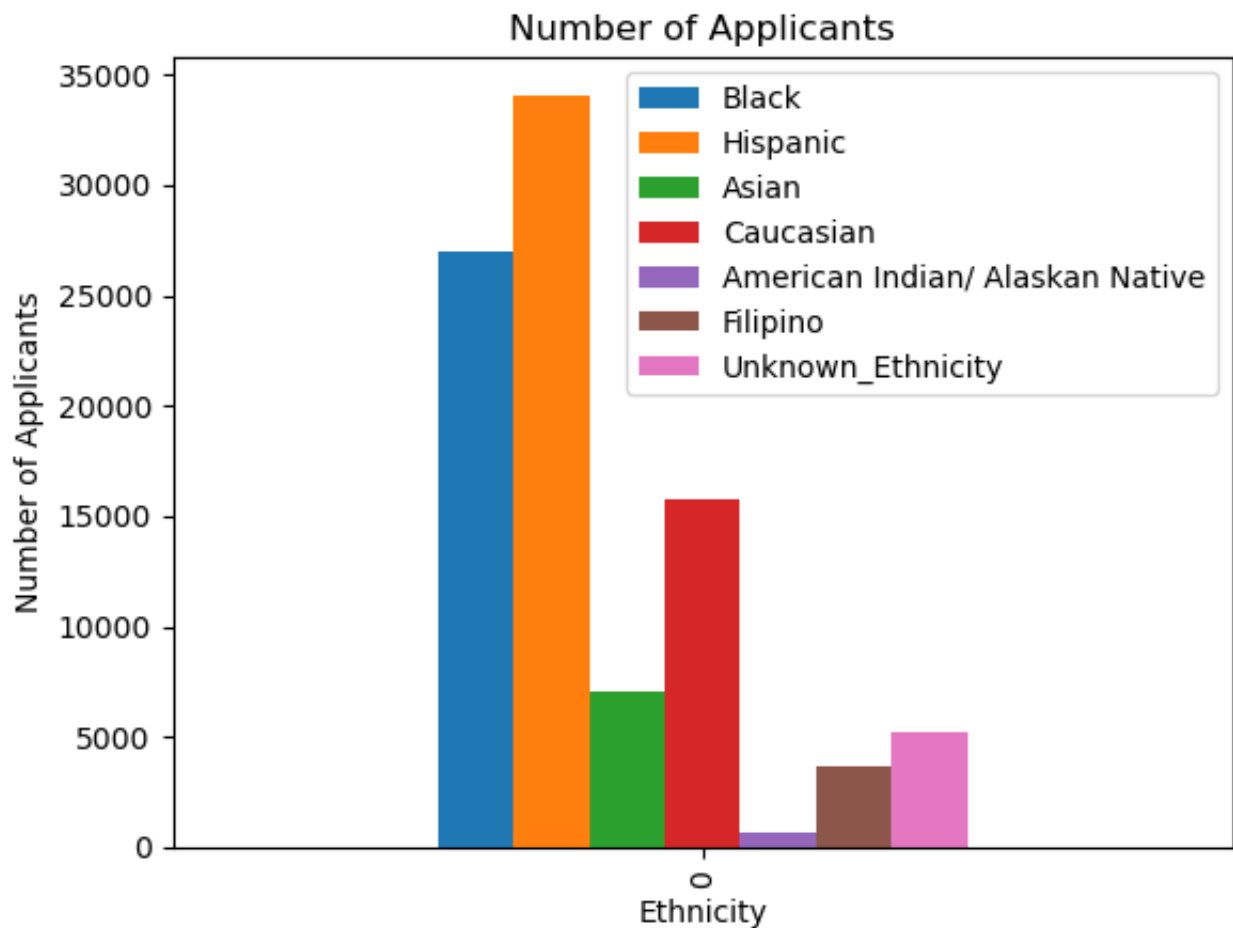
```

def ethnicity_total():
    ethnicity_total = job_applicants_df[
        [
            "Black",
            "Hispanic",
            "Asian",
            "Caucasian",
            "American Indian/ Alaskan Native",
            "Filipino",
            "Unknown_Ethnicity",
        ]
    ].sum()
    return ethnicity_total

```

```
# visualize the total number of applicants by ethnicity
def eth_chart():
    eth_and_total = ethnicity_total()
    eth_and_total = eth_and_total.to_pandas()
    eth_and_total.plot(kind="bar", stacked=False, title="Number of Applicant
    plt.xlabel("Ethnicity")
    plt.ylabel("Number of Applicants")
    plt.show()
    return eth_chart

ethnicity_total()
eth_chart()
```



Out[38]: <function __main__.eth_chart()>