Analyzing K-12 Student Data for School Year 2021

```
import numpy as np
import pandas as pd
import seaborn as sns
```

```
df = pd.read_csv("data-science-project-2021-k12.csv")
```

```
df
```

	SchoolType	EnrollmentType	SchoolCode	Track	Grade	GradeLevel	StudentID	HasLaptop	HasInternet
0	Traditional	Traditional	80	T	12	High	5	Yes	Yes
1	Traditional	Traditional	80	STE	11	High	7	Yes	Yes
2	Traditional	Traditional	80	T	12	High	13	Yes	Yes
3	Traditional	Traditional	80	T	12	High	2	No	No
4	Traditional	Traditional	80	T	11	High	15	Yes	Yes
...	...	...	...	...	...	...	...	...	...
51349	Traditional	Traditional	55	T	08	Intermediate	51338	No	No
51350	Year Round	Traditional	43	B	KA	Elementary	51343	No	Yes
51351	Traditional	Traditional	21	T	KN	Elementary	51351	Yes	Yes
51352	Traditional	Traditional	44	T	KN	Elementary	51346	No	No
51353	Year Round	Traditional	36	B	04	Elementary	51354	No	No

51354 rows × 34 columns

Name columns

```
df.columns
```

```
Index(['SchoolType', 'EnrollmentType', 'SchoolCode', 'Track', 'Grade',
       'GradeLevel', 'StudentID', 'HasLaptop', 'HasInternet',
       'LaptopCheckedOut', 'PCSurveyLaptop', 'MiFiCheckedOut',
       'PCSurveyInternet', 'PCSurveyInstrProg', 'TotalStudentInHousehold',
       'IsSED', 'IsSPED', 'SPEDProgram', 'IsEL', 'ELPAC', 'IsDLI', 'IsGATE',
       'IsGNP', 'IsFoster', 'Gender', 'Ethnicity', 'ParentHighestEdLevel',
       'DaysEnrolled', 'DaysPresent', 'PresentRate', 'CountT1P_N_IDF',
       'CountT1_N_DF', 'CountQ1P_IDF', 'CountQ1I_DF'],
      dtype='object')
```

Basic statistic

```
df.describe()
```

	SchoolCode	StudentID	TotalStudentInHousehold	ELPAC	DaysEnrolled	DaysPresent	Presen
count	51354.000000	51354.000000	51354.000000	5672.000000	51354.000000	51354.000000	51354.
mean	46.380788	25677.500000	1.775986	2.569817	32.782256	32.058652	0.9743
std	25.434033	14824.767199	1.413972	0.863895	5.676303	6.205820	0.0985
min	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.0000
25%	23.000000	12839.250000	1.000000	2.000000	32.000000	32.000000	1.0000
50%	48.000000	25677.500000	2.000000	3.000000	32.000000	32.000000	1.0000
75%	68.000000	38515.750000	2.000000	3.000000	32.000000	32.000000	1.0000
max	83.000000	51354.000000	41.000000	4.000000	57.000000	57.000000	1.0000

Make a copy of raw data before change data.

```
k_12 = df.copy()
```

Grade level distribution

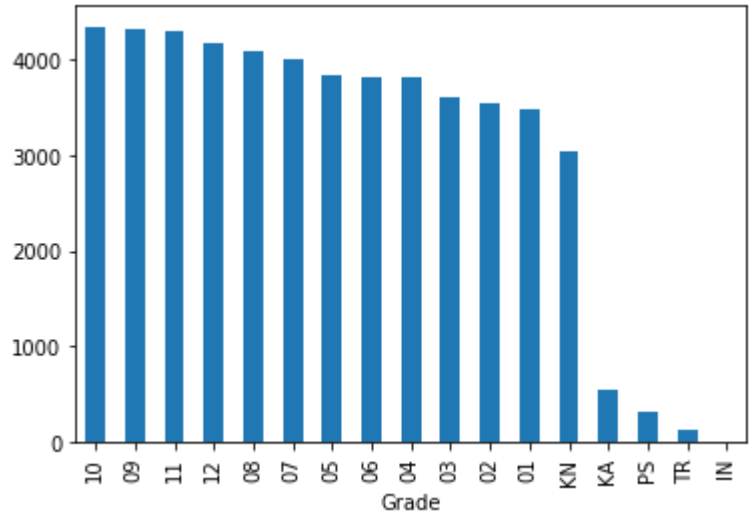
```
grade_level = k_12.groupby("Grade")["StudentID"].count().sort_values(ascending=Fa
grade_level
```

```
Grade
10    4350
09    4331
11    4298
12    4167
08    4084
07    4015
05    3832
06    3822
04    3813
03    3609
02    3552
01    3474
KN    3035
KA     538
PS     303
TR     130
IN       1
Name: StudentID, dtype: int64
```

```
grade_level.plot.bar()
```

```
<AxesSubplot:xlabel='Grade'>
```

Download



Ethnicity distribution

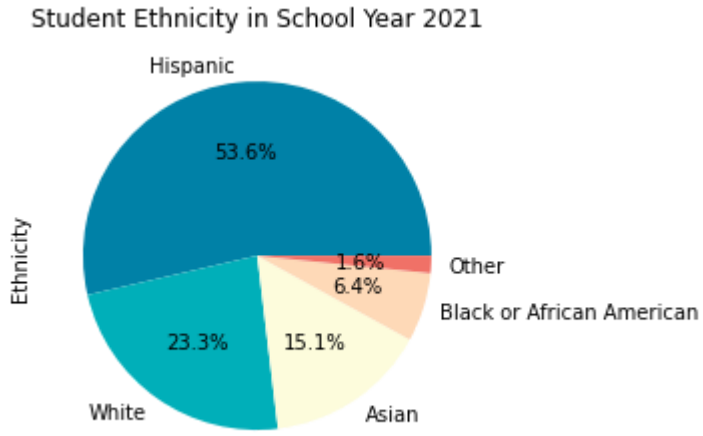
```
ethnicity_unique = k_12["Ethnicity"].value_counts()
ethnicity_unique
```

```
Hispanic                    27536
White                       11953
Asian                        7758
Black or African American    3284
Other                         823
Name: Ethnicity, dtype: int64
```

```
color_ethnicity = ["#0081a7","#00afb9","#fdfcdc","#fed9b7","#f07167"]
ethnicity_unique.plot.pie(colors = color_ethnicity,autopct='%1.1f%%', title = "St
```

```
<AxesSubplot:title={'center':'Student Ethnicity in School Year 2021'}, ylabel='
```

Download



Virtual vs Traditional Setting

Because of COVID-19, all California schools began school year 2021 in a remote learning (virtual) setting

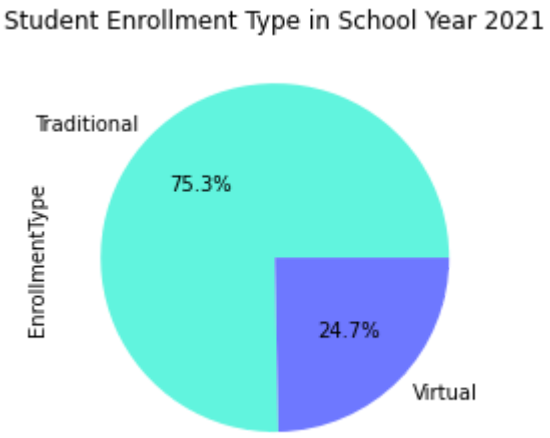
```
EnrollmentType_unique = k_12["EnrollmentType"].value_counts()
EnrollmentType_unique
```

```
Traditional    38678
Virtual        12676
Name: EnrollmentType, dtype: int64
```

```
EnrollmentType_unique.plot.pie(colors = ["#61f4de","#6e78ff"],autopct='%1.1f%%',
```

```
<AxesSubplot:title={'center':'Student Enrollment Type in School Year 2021'}, yl
```

Download



Device Checked Out Status

some families who can't afford these technologies, school district has budget to purchase laptop and internet for students to check out to use at home. Parents can make requests for laptop or/and internet on parent portal. Information Technology department gathers these requests and notifies parents to come to district office to pick up devices when inventory allows.

Has laptop

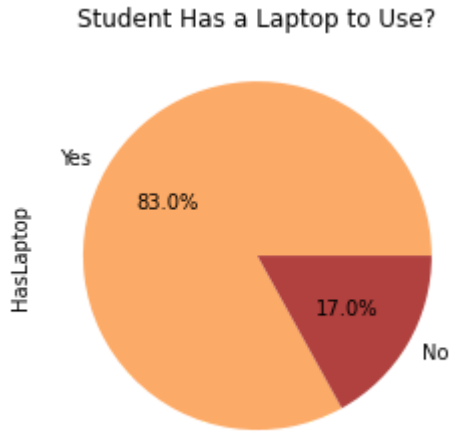
```
has_laptop_count = k_12["HasLaptop"].value_counts()
has_laptop_count
```

```
Yes    42649
No      8705
Name: HasLaptop, dtype: int64
```

```
has_laptop_count.plot.pie(colors = ["#fcaa67","#b0413e"],autopct='%1.1f%%', title
```

```
<AxesSubplot:title={'center':'Student Has a Laptop to Use?'}, ylabel='HasLaptop
```

Download



Has internet

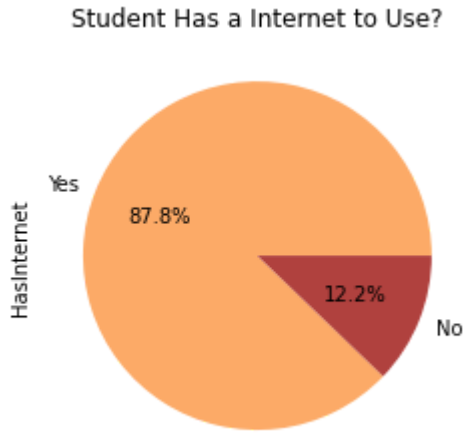
```
has_internet = k_12["HasInternet"].value_counts()
has_internet
```

```
Yes    45106
No      6248
Name: HasInternet, dtype: int64
```

```
has_internet.plot.pie(colors = ["#fcaa67","#b0413e"],autopct='%1.1f%%', title = "
```

```
<AxesSubplot:title={'center':'Student Has a Internet to Use?'}, ylabel='HasInte
```

Download



What ethnicity and/or gender group(s) of students prefer to start school year in a Virtual setting?

```
k_12["Count"] = k_12["StudentID"]
```

```
k_12
```

	SchoolType	EnrollmentType	SchoolCode	Track	Grade	GradeLevel	StudentID	HasLaptop	HasInternet
0	Traditional	Traditional	80	T	12	High	5	Yes	Yes
1	Traditional	Traditional	80	STE	11	High	7	Yes	Yes
2	Traditional	Traditional	80	T	12	High	13	Yes	Yes
3	Traditional	Traditional	80	T	12	High	2	No	No
4	Traditional	Traditional	80	T	11	High	15	Yes	Yes
...	...	...	...	...	...	...	...	...	...
51349	Traditional	Traditional	55	T	08	Intermediate	51338	No	No
51350	Year Round	Traditional	43	B	KA	Elementary	51343	No	Yes
51351	Traditional	Traditional	21	T	KN	Elementary	51351	Yes	Yes
51352	Traditional	Traditional	44	T	KN	Elementary	51346	No	No
51353	Year Round	Traditional	36	B	04	Elementary	51354	No	No

51354 rows × 35 columns

Virtual

```
virtual_k_12 = k_12[k_12["EnrollmentType"] == "Virtual"].groupby(["Ethnicity","En
virtual_k_12
```

	Ethnicity	EnrollmentType	Gender	Count
0	Asian	Virtual	Female	1338
1	Asian	Virtual	Male	1346
2	Black or African American	Virtual	Female	466
3	Black or African American	Virtual	Male	467
4	Hispanic	Virtual	Female	3516
5	Hispanic	Virtual	Male	3465
6	Hispanic	Virtual	Non-Binary	2
7	Other	Virtual	Female	91
8	Other	Virtual	Male	90
9	White	Virtual	Female	942
10	White	Virtual	Male	952
11	White	Virtual	Non-Binary	1

Traditional

```
traditional_k_12 = k_12[k_12["EnrollmentType"] == "Traditional"].groupby(["Ethnic
traditional_k_12
```

	Ethnicity	EnrollmentType	Gender	Count
0	Asian	Traditional	Female	2470
1	Asian	Traditional	Male	2604
2	Black or African American	Traditional	Female	1127
3	Black or African American	Traditional	Male	1224
4	Hispanic	Traditional	Female	10012
5	Hispanic	Traditional	Male	10539
6	Hispanic	Traditional	Non-Binary	2
7	Other	Traditional	Female	330
8	Other	Traditional	Male	311
9	Other	Traditional	Non-Binary	1
10	White	Traditional	Female	4849
11	White	Traditional	Male	5208
12	White	Traditional	Non-Binary	1

Merge

```
k_12_merge = virtual_k_12.merge(traditional_k_12, on = ["Ethnicity","Gender"],suf
k_12_merge
```

	Ethnicity	EnrollmentType_virtual	Gender	Count_virtual	EnrollmentType_traditional	Count_traditional
0	Asian	Virtual	Female	1338	Traditional	2470
1	Asian	Virtual	Male	1346	Traditional	2604
2	Black or African American	Virtual	Female	466	Traditional	1127
3	Black or African American	Virtual	Male	467	Traditional	1224
4	Hispanic	Virtual	Female	3516	Traditional	10012
5	Hispanic	Virtual	Male	3465	Traditional	10539
6	Hispanic	Virtual	Non-Binary	2	Traditional	2
7	Other	Virtual	Female	91	Traditional	330
8	Other	Virtual	Male	90	Traditional	311
9	White	Virtual	Female	942	Traditional	4849
10	White	Virtual	Male	952	Traditional	5208
11	White	Virtual	Non-Binary	1	Traditional	1

Virtual rate

```
k_12_merge["Virtual_rate"] = k_12_merge["Count_virtual"] / (k_12_merge["Count_vir
k_12_merge
```

	Ethnicity	EnrollmentType_virtual	Gender	Count_virtual	EnrollmentType_traditional	Count_traditional	Virtual_
0	Asian	Virtual	Female	1338	Traditional	2470	0.3513
1	Asian	Virtual	Male	1346	Traditional	2604	0.3407
2	Black or African American	Virtual	Female	466	Traditional	1127	0.2925
3	Black or African American	Virtual	Male	467	Traditional	1224	0.2761
4	Hispanic	Virtual	Female	3516	Traditional	10012	0.2599
5	Hispanic	Virtual	Male	3465	Traditional	10539	0.2474
6	Hispanic	Virtual	Non-Binary	2	Traditional	2	0.5000
7	Other	Virtual	Female	91	Traditional	330	0.2161
8	Other	Virtual	Male	90	Traditional	311	0.2244
9	White	Virtual	Female	942	Traditional	4849	0.1626
10	White	Virtual	Male	952	Traditional	5208	0.1545
11	White	Virtual	Non-Binary	1	Traditional	1	0.5000

```
k_12_merge.columns
```

```
Index(['Ethnicity', 'EnrollmentType_virtual', 'Gender', 'Count_virtual',
       'EnrollmentType_traditional', 'Count_traditional', 'Virtual_rate'],
      dtype='object')
```

```
k_12_merge_pivot = k_12_merge.pivot("Ethnicity","Gender","Virtual_rate").fillna(0
k_12_merge_pivot
```

Gender	Female	Male	Non-Binary
Ethnicity			
Asian	0.351366	0.340759	0.0
Black or African American	0.292530	0.276168	0.0
Hispanic	0.259905	0.247429	0.5
Other	0.216152	0.224439	0.0
White	0.162666	0.154545	0.5

```
colormap = sns.color_palette("crest", as_cmap=True)
sns.heatmap(k_12_merge_pivot, annot=True, cmap=colormap)
```

```
<AxesSubplot:xlabel='Gender', ylabel='Ethnicity'>
```

Download