

prelim_analysis_omar

October 25, 2024

```
[1]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
[2]: cancelled = pd.read_csv('cancelled.csv')
cancelled
```

```
[2]:
```

	OptionsNumber	StartDate	Duration	AppointmentStatus	\
0	89175	7/1/21 1:00	61	No Show/Client	
1	61751	7/1/21 8:00	90	Provider Cx	
2	62043	7/1/21 8:45	120	No Show/Client	
3	57251	7/1/21 9:00	120	Other Agency Cxl	
4	60712	7/1/21 9:00	60	Provider Cx	
...	
52628	91696	6/28/24 16:15	0	Client Cx	
52629	90917	6/28/24 16:30	60	No Show/Client	
52630	68578	6/28/24 17:30	0	Client Cx	
52631	89046	6/28/24 17:30	60	Provider Cx	
52632	65318	6/30/24 9:00	60	No Show/Client	

	AgeAtAppointment	Gender	Ethnicity	PrimaryLanguage	\
0	17	M	Black	English	
1	27	F	Hispanic/Latino	Spanish	
2	17	F	White	English	
3	16	M	Hispanic/Latino	English	
4	1	M	Hispanic/Latino	English	
...	
52628	8	F	White	English	
52629	15	F	Hispanic/Latino	English	
52630	15	F	Hispanic/Latino	English	
52631	8	M	White	English	
52632	16	M	Hispanic/Latino	English	

	Religion	County
0	[Not Entered]	Los Angeles
1	[Not Entered]	Ventura
2	[Not Entered]	Stanislaus

3	[Not Entered]	Merced
4	[Not Entered]	Stanislaus
...
52628	[Not Entered]	Kern
52629	[Not Entered]	Tulare
52630	[Not Entered]	Ventura
52631	[Not Entered]	Stanislaus
52632	[Not Entered]	Merced

[52633 rows x 10 columns]

```
[3]: incidents = pd.read_csv('incident.csv')
incidents
```

```
[3]:      IncidentDate  IncidentTypeCode  \
0      02/01/2022                6
1      11/26/2022                6
2      11/26/2022                6
3      02/01/2022                6
4      07/06/2022                2
...
6488    01/25/2023                20
6489    01/24/2023                18
6490    01/26/2023                55
6491    02/09/2023                58
6492    02/22/2023                27
```

	IncidentType	SubTypeCode	\
0	Epidemic Outbreak	NaN	
1	Epidemic Outbreak	NaN	
2	Epidemic Outbreak	NaN	
3	Epidemic Outbreak	NaN	
4	Unusual Incident	5.0	
...	
6488	12. Suicide Talk - screening	NaN	
6489	11. Suspected Child Abuse (CPS) Report	NaN	
6490	Suicide	132.0	
6491	Unusual Incident	149.0	
6492	06. Infection Control/Epidemic Outbreak Issue	NaN	

	IncidentSubType	OptionsNumber	AgeAtIncident	Gender	Ethnicity	\
0	NaN	77036	13	F	White	
1	NaN	77036	13	F	White	
2	NaN	77037	15	F	White	
3	NaN	77037	14	F	White	
4	ER Visit	78745	13	F	Hispanic/Latino	
...	

6488	NaN	90893	10	M	White
6489	NaN	90826	17	F	White
6490	Talk	100019	19	U	[Not Entered]
6491	Other	100024	19	F	Black
6492	NaN	65185	12	F	Hispanic/Latino

	PrimaryLanguage	Religion	ServiceDivision \
0	English	Non-Denominational	Foster Care
1	English	Non-Denominational	Foster Care
2	English	No Preference	Foster Care
3	English	No Preference	Foster Care
4	English	No Preference	Foster Care
...
6488	English	[Not Entered]	Wraparound
6489	English	[Not Entered]	Wraparound
6490	[Not Entered]	[Not Entered]	Transitional
6491	English	No Preference	Transitional
6492	English	[Not Entered]	Behavioral Health

	County	Service
0	FRESNO	DSS
1	FRESNO	DSS
2	FRESNO	DSS
3	FRESNO	DSS
4	FRESNO	DSS
...
6488	WRAP - Kern	Kaiser
6489	WRAP - Tulare	WRAP Mental Health
6490	L.A.	DSS
6491	STANISLAUS	DSS
6492	BH-Merced County	CSP

[6493 rows x 13 columns]

```
[4]: challenges = pd.read_csv('challenges.csv')
challenges
```

	OptionsNumber	DateCompleted	CANSCategoryName \
0	51302	11/02/2019	Caregiver Strengths & Needs
1	51302	11/02/2019	Caregiver Strengths & Needs
2	51302	11/02/2019	Caregiver Strengths & Needs
3	51302	11/02/2019	Caregiver Strengths & Needs
4	51302	11/02/2019	Caregiver Strengths & Needs
...
597761	69680	09/18/2024	x-Module: Trauma
597762	69680	09/18/2024	x-Module: Trauma
597763	69680	09/18/2024	x-Module: Trauma

597764	69680	09/18/2024	x-Module: Trauma
597765	69680	09/18/2024	x-Module: Trauma

	ItemTitle	QuestionKey	\
0	Accessibility To Child Care Services	0-3	
1	Developmental	0-3	
2	Educational Attainment	0-3	
3	Employment/Educational Functioning	0-3	
4	Family Stress	0-3	
...	
597761	Sexual Abuse (Y/N)	Y/N	
597762	War/Terroism Affected (Y/N)	Y/N	
597763	Witness to Community or School Violence (Y/N)	Y/N	
597764	Witness to Family Violence (Y/N)	Y/N	
597765	Witness/Victim to Criminal Activity (Y/N)	Y/N	

	ChoiceValue	AgeWhenAssessed	Gender	Ethnicity	PrimaryLanguage	\
0	0.0	16	F	Laotian	English	
1	0.0	16	F	Laotian	English	
2	0.0	16	F	Laotian	English	
3	0.0	16	F	Laotian	English	
4	0.0	16	F	Laotian	English	
...	
597761	0.0	6	F	No Preference	English	
597762	0.0	6	F	No Preference	English	
597763	1.0	6	F	No Preference	English	
597764	1.0	6	F	No Preference	English	
597765	0.0	6	F	No Preference	English	

	Religion	County
0	[Not Entered]	Merced
1	[Not Entered]	Merced
2	[Not Entered]	Merced
3	[Not Entered]	Merced
4	[Not Entered]	Merced
...
597761	[Not Entered]	Stanislaus
597762	[Not Entered]	Stanislaus
597763	[Not Entered]	Stanislaus
597764	[Not Entered]	Stanislaus
597765	[Not Entered]	Stanislaus

[597766 rows x 12 columns]

```
[5]: demographic = pd.read_csv('demographic.csv')
demographic
```

```

[5]: OptionsNumber Gender AgeAtStart Ethnicity PrimaryLanguage \
0      8425      F      5      [Not Entered]      [Not Entered]
1     13081      M     11      [Not Entered]      [Not Entered]
2     14959      F     16      Hispanic/Latino      English
3     15113      F     15      [Not Entered]      English
4     16063      F     22           Black      English
...
4445     100046      F     22      No Preference      English
4446     100048      F     18      Hispanic/Latino      English
4447     100050      U     19      [Not Entered]      [Not Entered]
4448     100051      U     18      [Not Entered]      [Not Entered]
4449     100052      U     18      [Not Entered]      [Not Entered]

      Religion      ServiceDivision      CountyofService \
0      [Not Entered]      Behavioral Health      ZBH-Stan Co. Outpatient (Do not use)
1      [Not Entered]      Behavioral Health      ZBH-Stan Co. Outpatient (Do not use)
2      [Not Entered]      Behavioral Health      ZBH-Stan Co. Outpatient (Do not use)
3      [Not Entered]      Behavioral Health      ZBH-Stan Co. Outpatient (Do not use)
4      No Preference      Transitional      SAN JOAQUIN DSS
...
4445      Other Religion      Transitional      SAN JOAQUIN DSS
4446      No Preference      Transitional      KERN DSS
4447      [Not Entered]      Transitional      L.A. DSS
4448      [Not Entered]      Transitional      L.A. DSS
4449      [Not Entered]      Transitional      RIVERSIDE DSS

      DOP      DOD
0      07/21/2010      NaN
1      09/13/2011      NaN
2      11/08/2013      NaN
3      12/04/2013      NaN
4      09/02/2022      NaN
...
4445      01/30/2023      NaN
4446      01/31/2023      NaN
4447      02/16/2023      NaN
4448      02/22/2023      NaN
4449      02/22/2023      NaN

[4450 rows x 10 columns]

```

```

[6]: cancelled.sort_values(by='OptionsNumber', ascending=True)

```

```

[6]: OptionsNumber      StartDate      Duration      AppointmentStatus \
21767      8107      9/12/22 15:00      60      Client Cx
6073      8107      11/2/21 16:00      60      Other Agency Cx1
3486      11745      9/8/21 16:00      60      Other Agency Cx1

```

22142	11745	9/20/22	13:30	60	Provider Cx
12193	11745	3/14/22	16:30	60	Client Cx
...
33940	92521	7/7/23	10:00	60	Client Cx
5118	92521	10/13/21	14:00	60	No Show/Client
12273	92521	3/15/22	16:00	60	No Show/Client
10313	92521	2/2/22	16:30	60	No Show/Client
42252	92521	1/9/24	15:30	60	No Show/Client

	AgeAtAppointment	Gender	Ethnicity	PrimaryLanguage	\
21767	18	F	White	English	
6073	17	F	White	English	
3486	15	M	White	English	
22142	17	M	White	English	
12193	16	M	White	English	
...
33940	17	F	Hispanic/Latino	English	
5118	15	F	Hispanic/Latino	English	
12273	15	F	Hispanic/Latino	English	
10313	15	F	Hispanic/Latino	English	
42252	17	F	Hispanic/Latino	English	

	Religion	County
21767	[Not Entered]	Stanislaus
6073	[Not Entered]	Stanislaus
3486	[Not Entered]	Stanislaus
22142	[Not Entered]	Stanislaus
12193	[Not Entered]	Stanislaus
...
33940	[Not Entered]	Stanislaus
5118	[Not Entered]	Stanislaus
12273	[Not Entered]	Stanislaus
10313	[Not Entered]	Stanislaus
42252	[Not Entered]	Stanislaus

[52633 rows x 10 columns]

```
[7]: cancelled['AppointmentStatus'].unique() # What do these all mean? Are these
      ↪ irrelevant? Should we filter on any of these?
```

```
[7]: array(['No Show/Client', 'Provider Cx', 'Other Agency Cx1', 'Late Cx1',
      'Client Cx', 'County Disallowed'], dtype=object)
```

```
[8]: cancelled[cancelled['AppointmentStatus'] == 'County Disallowed']
```

```
[8]: OptionsNumber      StartDate  Duration  AppointmentStatus \
      184             58196    7/7/21 9:00          30  County Disallowed
```

539	54969	7/13/21	11:00	105	County Disallowed
1274	89563	7/26/21	17:45	19	County Disallowed
1413	89563	7/28/21	17:45	133	County Disallowed
1490	89563	7/29/21	17:45	123	County Disallowed
...
32544	64186	6/5/23	11:30	120	County Disallowed
32600	63254	6/6/23	11:45	3	County Disallowed
32642	65299	6/6/23	18:20	60	County Disallowed
32686	65849	6/7/23	14:30	95	County Disallowed
40349	65766	12/1/23	11:00	70	County Disallowed

	AgeAtAppointment	Gender	Ethnicity	PrimaryLanguage	\
184	8	M	White	English	
539	15	M	Hispanic/Latino	English	
1274	12	F	Mexican	English	
1413	12	F	Mexican	English	
1490	12	F	Mexican	English	
...
32544	18	F	Hispanic/Latino	English	
32600	8	F	Black	English	
32642	13	F	White	English	
32686	7	F	Hispanic/Latino	English	
40349	12	F	Hispanic/Latino	English	

	Religion	County
184	[Not Entered]	Stanislaus
539	[Not Entered]	San Benito
1274	[Not Entered]	Stanislaus
1413	[Not Entered]	Stanislaus
1490	[Not Entered]	Stanislaus
...
32544	[Not Entered]	San Joaquin
32600	[Not Entered]	San Joaquin
32642	[Not Entered]	San Joaquin
32686	[Not Entered]	San Joaquin
40349	[Not Entered]	Stanislaus

[200 rows x 10 columns]

```
[9]: cancelled.sort_values(by='Duration', ascending=True)
```

	OptionsNumber	StartDate	Duration	AppointmentStatus	\
39276	20007	11/6/23 17:15	0	Late Cxl	
24204	90498	11/2/22 16:00	0	Client Cx	
24205	90498	11/2/22 16:00	0	No Show/Client	
12267	89441	3/15/22 15:30	0	Client Cx	
29387	91007	3/30/23 12:30	0	Client Cx	

...
31945	91041	5/20/23	8:00	360	Client Cx
15458	61152	5/16/22	9:15	360	No Show/Client
4966	61549	10/11/21	9:30	480	Provider Cx
30204	60286	4/18/23	9:00	540	Provider Cx
4440	62342	9/28/21	12:30	750	No Show/Client

	AgeAtAppointment	Gender	Ethnicity	PrimaryLanguage	\
39276		7 F	Other Asian	English	
24204		17 F	No Preference	English	
24205		17 F	No Preference	English	
12267		18 M	White	English	
29387		18 M	No Preference	English	

...
31945		13 F	Hispanic/Latino	English	
15458		13 F	Hispanic/Latino	English	
4966		13 F	White	English	
30204		16 F	Black	English	
4440		13 F	Black	English	

	Religion	County
39276	[Not Entered]	Stanislaus
24204	[Not Entered]	Kern
24205	[Not Entered]	Kern
12267	[Not Entered]	Merced
29387	[Not Entered]	Santa Clara
...
31945	[Not Entered]	Kern
15458	[Not Entered]	Merced
4966	[Not Entered]	Merced
30204	[Not Entered]	Kings
4440	[Not Entered]	San Joaquin

[52633 rows x 10 columns]

```
[10]: cancelled[cancelled['Duration'] == 0]
```

```
[10]:
```

	OptionsNumber	StartDate	Duration	AppointmentStatus	\
5	60942	7/1/21 9:00	0	Late Cxl	
26	89055	7/1/21 12:15	0	Client Cx	
27	89206	7/1/21 12:57	0	No Show/Client	
40	76465	7/1/21 14:00	0	Client Cx	
45	89065	7/1/21 14:30	0	Client Cx	
...	
52620	68916	6/28/24 15:30	0	Client Cx	
52621	92211	6/28/24 15:30	0	Client Cx	
52623	91904	6/28/24 15:58	0	No Show/Client	

52628	91696	6/28/24 16:15	0	Client Cx
52630	68578	6/28/24 17:30	0	Client Cx

	AgeAtAppointment	Gender	Ethnicity	PrimaryLanguage	\
5	5	F	Mexican	English	
26	10	M	White	English	
27	17	F	White	English	
40	15	M	Hispanic/Latino	English	
45	15	F	Hispanic/Latino	English	
...	
52620	10	M	Hispanic/Latino	English	
52621	8	M	Black	English	
52623	14	F	Hispanic/Latino	English	
52628	8	F	White	English	
52630	15	F	Hispanic/Latino	English	

	Religion	County
5	[Not Entered]	Ventura
26	[Not Entered]	Stanislaus
27	[Not Entered]	Kern
40	[Not Entered]	Merced
45	[Not Entered]	Merced
...
52620	[Not Entered]	Merced
52621	[Not Entered]	Kern
52623	[Not Entered]	Tulare
52628	[Not Entered]	Kern
52630	[Not Entered]	Ventura

[8490 rows x 10 columns]

[11]: cancelled

	OptionsNumber	StartDate	Duration	AppointmentStatus	\
0	89175	7/1/21 1:00	61	No Show/Client	
1	61751	7/1/21 8:00	90	Provider Cx	
2	62043	7/1/21 8:45	120	No Show/Client	
3	57251	7/1/21 9:00	120	Other Agency Cxl	
4	60712	7/1/21 9:00	60	Provider Cx	
...	
52628	91696	6/28/24 16:15	0	Client Cx	
52629	90917	6/28/24 16:30	60	No Show/Client	
52630	68578	6/28/24 17:30	0	Client Cx	
52631	89046	6/28/24 17:30	60	Provider Cx	
52632	65318	6/30/24 9:00	60	No Show/Client	

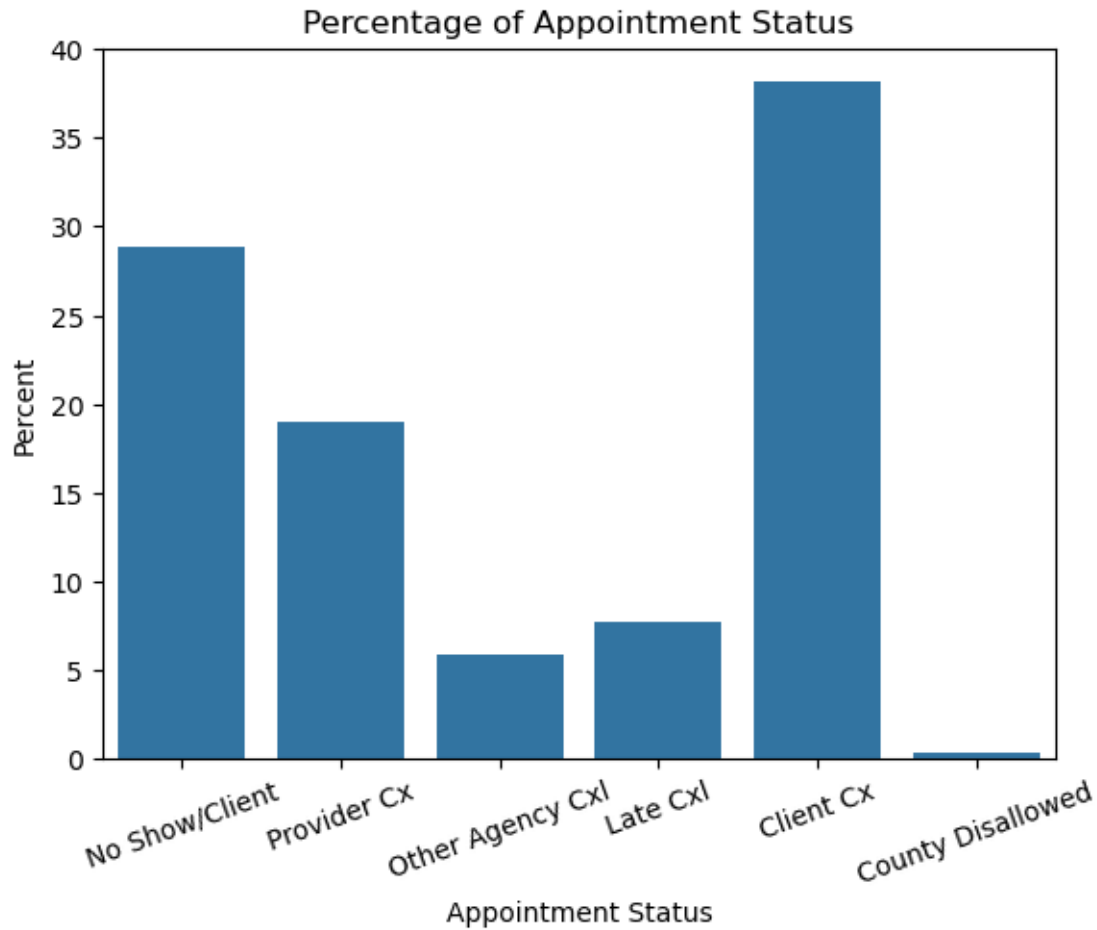
AgeAtAppointment	Gender	Ethnicity	PrimaryLanguage	\
------------------	--------	-----------	-----------------	---

0		17	M	Black	English
1		27	F	Hispanic/Latino	Spanish
2		17	F	White	English
3		16	M	Hispanic/Latino	English
4		1	M	Hispanic/Latino	English
...	
52628		8	F	White	English
52629		15	F	Hispanic/Latino	English
52630		15	F	Hispanic/Latino	English
52631		8	M	White	English
52632		16	M	Hispanic/Latino	English

	Religion	County
0	[Not Entered]	Los Angeles
1	[Not Entered]	Ventura
2	[Not Entered]	Stanislaus
3	[Not Entered]	Merced
4	[Not Entered]	Stanislaus
...
52628	[Not Entered]	Kern
52629	[Not Entered]	Tulare
52630	[Not Entered]	Ventura
52631	[Not Entered]	Stanislaus
52632	[Not Entered]	Merced

[52633 rows x 10 columns]

```
[12]: sns.countplot(data=cancelled, x='AppointmentStatus', stat='percent')
plt.xticks(rotation=20);
plt.xlabel('Appointment Status')
plt.ylabel('Percent')
plt.title('Percentage of Appointment Status');
```



```
[13]: no_shows = cancelled[cancelled['AppointmentStatus'] == 'No Show/Client']
      no_shows
```

```
[13]:
```

	OptionsNumber	StartDate	Duration	AppointmentStatus	\
0	89175	7/1/21 1:00	61	No Show/Client	
2	62043	7/1/21 8:45	120	No Show/Client	
7	53563	7/1/21 9:30	5	No Show/Client	
13	61848	7/1/21 10:00	120	No Show/Client	
14	55157	7/1/21 11:00	90	No Show/Client	
...	
52617	61138	6/28/24 15:30	90	No Show/Client	
52623	91904	6/28/24 15:58	0	No Show/Client	
52626	68003	6/28/24 16:00	60	No Show/Client	
52629	90917	6/28/24 16:30	60	No Show/Client	
52632	65318	6/30/24 9:00	60	No Show/Client	

	AgeAtAppointment	Gender	Ethnicity	PrimaryLanguage	\
0	17	M	Black	English	

2		17	F	White	English
7		17	F	White	English
13		17	F	Hispanic/Latino	English
14		10	M	No Preference	English
...
52617		16	F	Hispanic/Latino	English
52623		14	F	Hispanic/Latino	English
52626		11	M	Hispanic/Latino	English
52629		15	F	Hispanic/Latino	English
52632		16	M	Hispanic/Latino	English

	Religion	County
0	[Not Entered]	Los Angeles
2	[Not Entered]	Stanislaus
7	[Not Entered]	Stanislaus
13	[Not Entered]	Stanislaus
14	[Not Entered]	Merced
...
52617	[Not Entered]	Kings
52623	[Not Entered]	Tulare
52626	[Not Entered]	Stanislaus
52629	[Not Entered]	Tulare
52632	[Not Entered]	Merced

[15178 rows x 10 columns]

```
[14]: no_shows_counts = no_shows['OptionsNumber'].value_counts().to_frame()
      no_shows_counts
```

```
[14]:
```

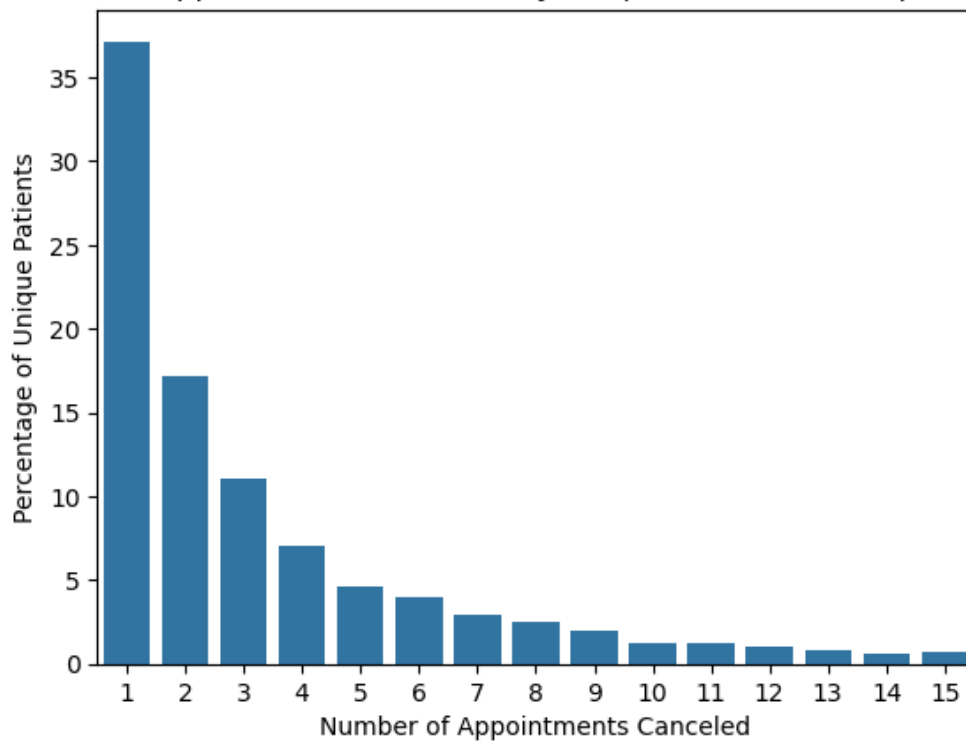
	count
OptionsNumber	
61138	115
62562	87
89244	85
89492	76
66291	70
...	...
67501	1
67673	1
68757	1
67263	1
67261	1

[3217 rows x 1 columns]

```
[15]: # here we can see that the vast majority of people are cancelling appointments
      ↪are typically just a one time thing
```

```
# however, we want to consider the extremely long tail of many cancelled
↳appointments
# i limited the x axis, because there is one individual who had 115 missed
↳appointments
# also for data visualization purposes to make it more readable
sns.countplot(data=no_shows_counts, x='count', stat='percent')
plt.ylabel('Percentage of Unique Patients')
plt.xlabel('Number of Appointments Canceled')
plt.title('Number of Appointments Canceled by Unique Patients in the past 3
↳years')
plt.xlim(-0.5, 14.5);
```

Number of Appointments Canceled by Unique Patients in the past 3 years



0.1 Cancelled Demographics

[16]: cancelled

```
[16]:
```

	OptionsNumber	StartDate	Duration	AppointmentStatus	\
0	89175	7/1/21 1:00	61	No Show/Client	
1	61751	7/1/21 8:00	90	Provider Cx	
2	62043	7/1/21 8:45	120	No Show/Client	
3	57251	7/1/21 9:00	120	Other Agency Cxl	

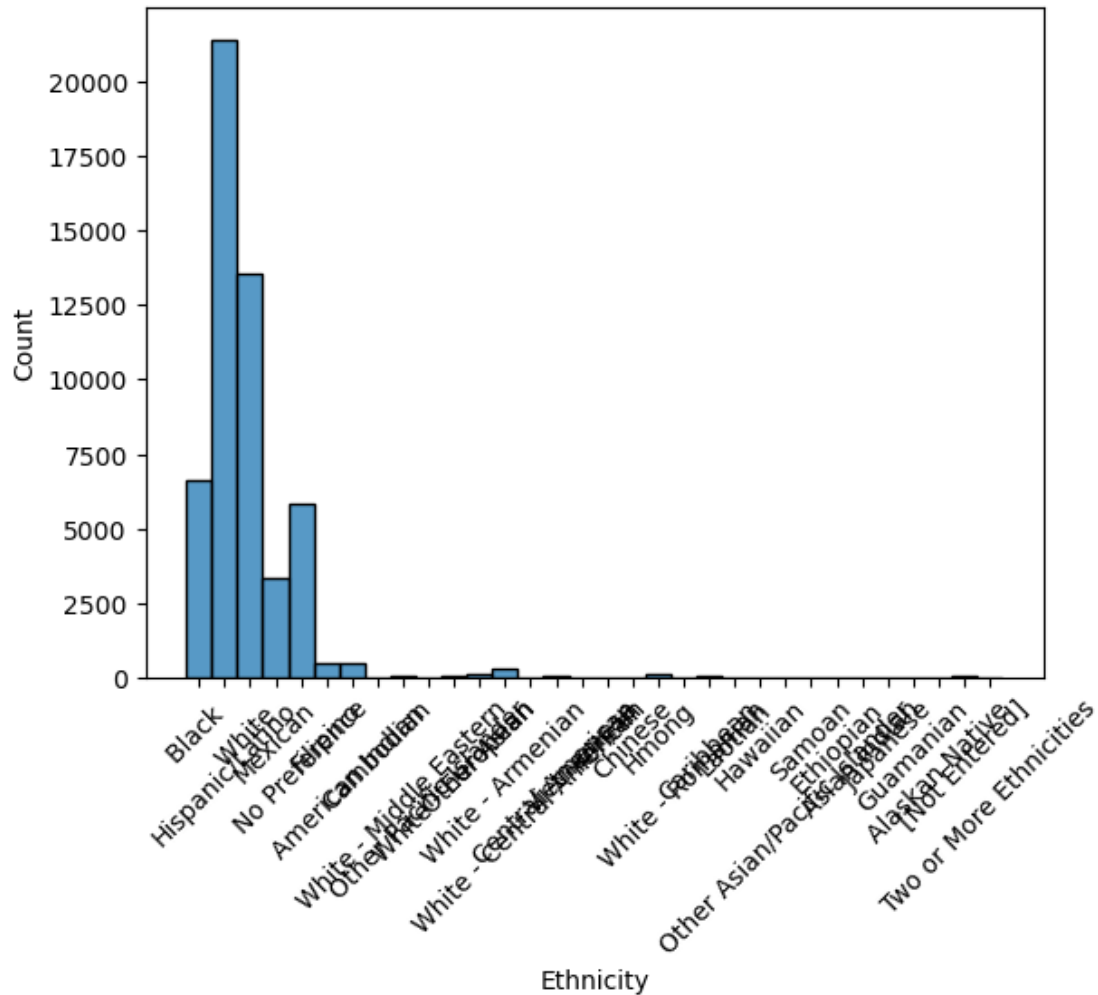
4	60712	7/1/21 9:00	60	Provider Cx
...
52628	91696	6/28/24 16:15	0	Client Cx
52629	90917	6/28/24 16:30	60	No Show/Client
52630	68578	6/28/24 17:30	0	Client Cx
52631	89046	6/28/24 17:30	60	Provider Cx
52632	65318	6/30/24 9:00	60	No Show/Client

	AgeAtAppointment	Gender	Ethnicity	PrimaryLanguage	\
0	17	M	Black	English	
1	27	F	Hispanic/Latino	Spanish	
2	17	F	White	English	
3	16	M	Hispanic/Latino	English	
4	1	M	Hispanic/Latino	English	
...	
52628	8	F	White	English	
52629	15	F	Hispanic/Latino	English	
52630	15	F	Hispanic/Latino	English	
52631	8	M	White	English	
52632	16	M	Hispanic/Latino	English	

	Religion	County
0	[Not Entered]	Los Angeles
1	[Not Entered]	Ventura
2	[Not Entered]	Stanislaus
3	[Not Entered]	Merced
4	[Not Entered]	Stanislaus
...
52628	[Not Entered]	Kern
52629	[Not Entered]	Tulare
52630	[Not Entered]	Ventura
52631	[Not Entered]	Stanislaus
52632	[Not Entered]	Merced

[52633 rows x 10 columns]

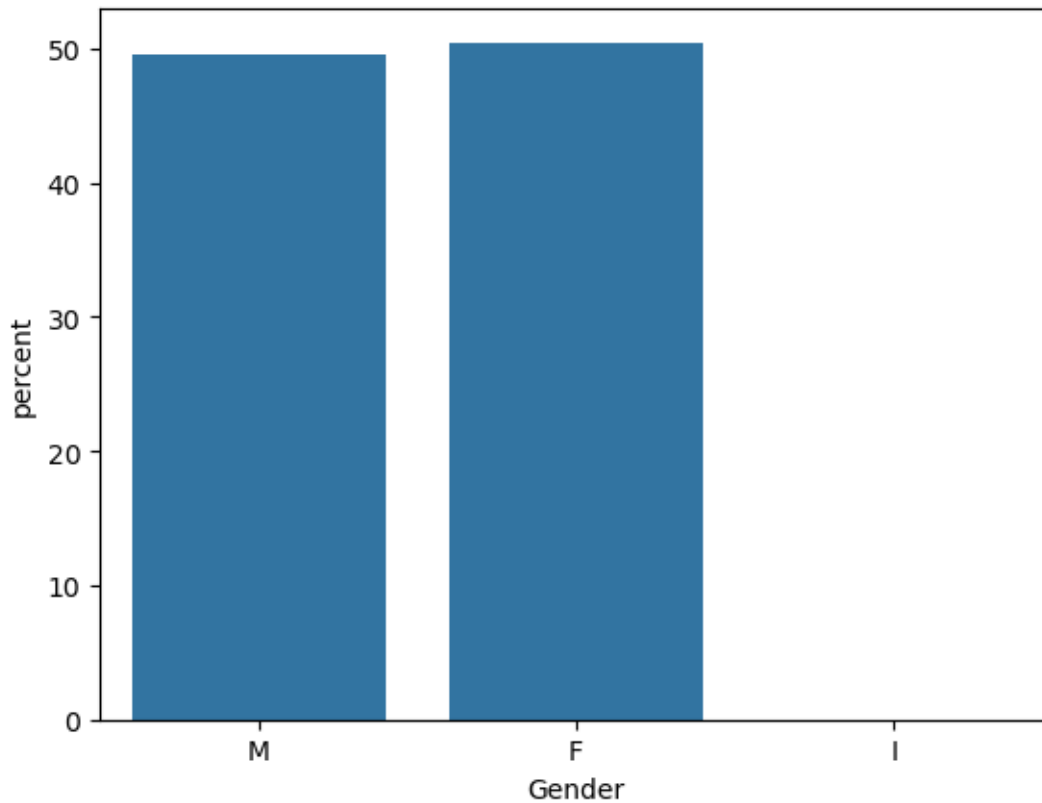
```
[17]: # Ethnicity
sns.histplot(data=cancelled, x='Ethnicity')
plt.xticks(rotation=45);
# eventually bin these ethnicities to make the graph look more cleaner (maybe
↳ like the top 5 + others?
```



```
[18]: cancelled['Ethnicity'].unique()
```

```
[18]: array(['Black ', 'Hispanic/Latino', 'White ', 'Mexican', 'No Preference',
        'Filipino', 'American Indian ', 'Cambodian',
        'White - Middle Eastern', 'Other Pacific Islander',
        'White - European', 'Other Asian', 'White - Armenian',
        'White - Central American', 'Central American', 'Vietnamese',
        'Korean', 'Chinese', 'Hmong', 'White - Romanian', 'Caribbean',
        'Laotian', 'Hawaiian', 'Other Asian/Pacific Islander', 'Samoan',
        'Ethiopian', 'Asian Indian', 'Japanese', 'Guamanian',
        'Alaskan Native', '[Not Entered]', 'Two or More Ethnicities'],
      dtype=object)
```

```
[19]: # Gender
      sns.countplot(data=cancelled, x='Gender', stat='percent');
```



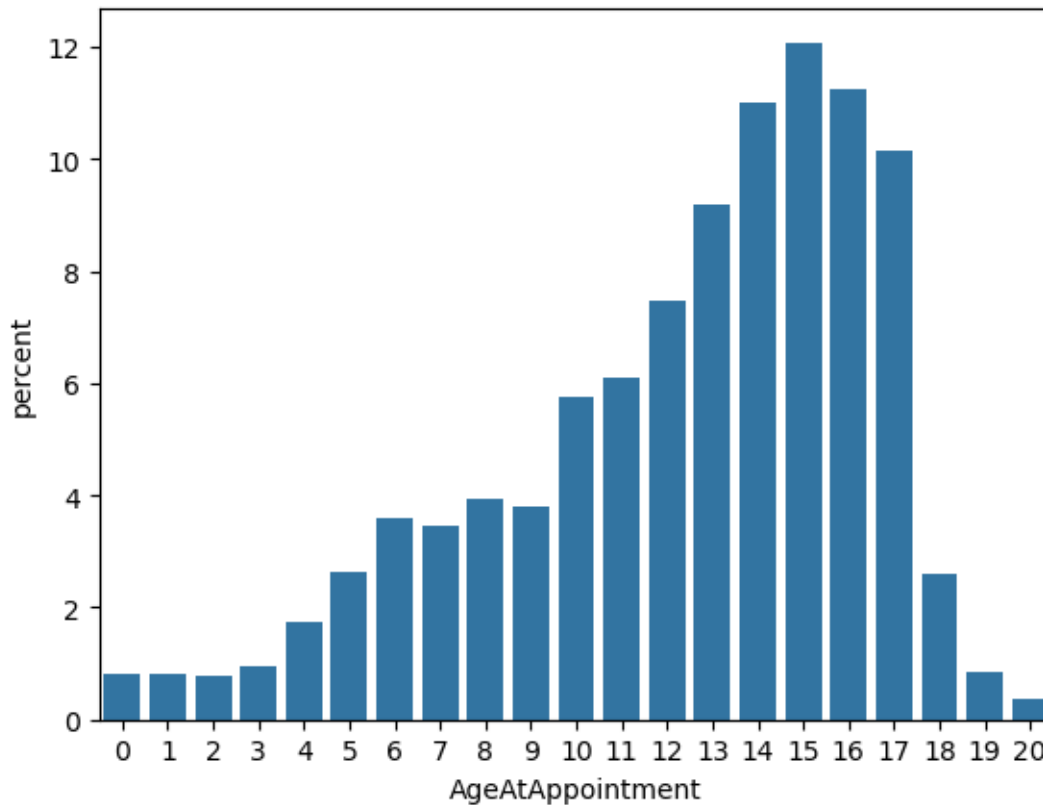
```
[20]: cancelled[cancelled['Gender'] == 'I']
# seems like there is a (singular) person with the gender I? maybe gender
↳ confirming or just not put?
```

```
[20]: OptionsNumber      StartDate      Duration AppointmentStatus \
16437          62795  6/1/22 18:00          90          Late Cxl

      AgeAtAppointment Gender      Ethnicity PrimaryLanguage      Religion \
16437          37      I  No Preference          English  [Not Entered]

      County
16437  Ventura
```

```
[21]: sns.countplot(data=cancelled, x='AgeAtAppointment', stat='percent')
plt.xlim(-0.5, 20.5);
```

```
[22]: # next steps:
      # turn time date into days of the week and see if there are any trends on what
      # day people cancelled on
      # same with months
```

```
[23]: cancelled[['Date', 'Time']] = cancelled['StartDate'].str.split(' ',
      # expand=True) #prepare for datetime change
cancelled['date'] = pd.to_datetime(cancelled['Date']) # object to dataframe
      # (can produce better data)
cancelled = cancelled.drop(['StartDate', 'Date'], axis=1)
cancelled['month'] = cancelled['date'].dt.month
cancelled['dotw'] = cancelled['date'].dt.day_name()
cancelled['int_dotw'] = cancelled['date'].dt.dayofweek
```

/tmp/ipykernel_111/440367340.py:2: UserWarning: Could not infer format, so each element will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.

```
cancelled['date'] = pd.to_datetime(cancelled['Date']) # object to dataframe
(can produce better data)
```

```
[24]: cancelled
```

```
[24]:
```

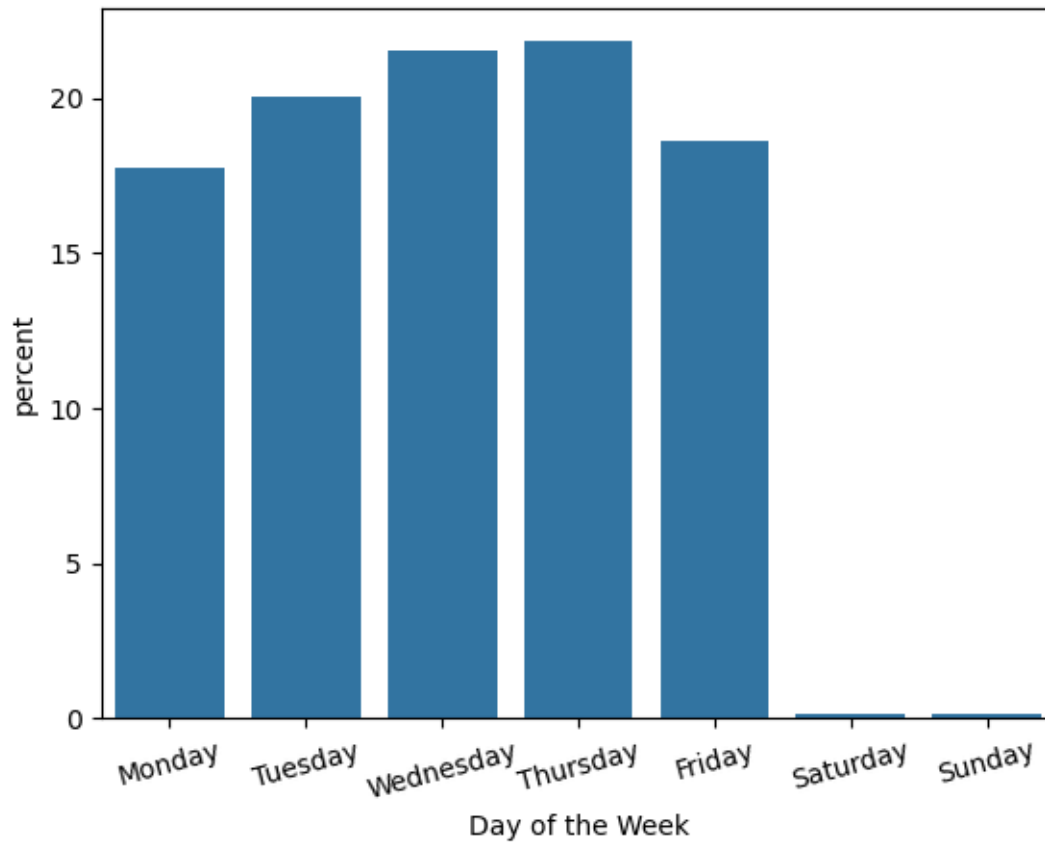
	OptionsNumber	Duration	AppointmentStatus	AgeAtAppointment	Gender	\
0	89175	61	No Show/Client	17	M	
1	61751	90	Provider Cx	27	F	
2	62043	120	No Show/Client	17	F	
3	57251	120	Other Agency Cxl	16	M	
4	60712	60	Provider Cx	1	M	
...	
52628	91696	0	Client Cx	8	F	
52629	90917	60	No Show/Client	15	F	
52630	68578	0	Client Cx	15	F	
52631	89046	60	Provider Cx	8	M	
52632	65318	60	No Show/Client	16	M	

	Ethnicity	PrimaryLanguage	Religion	County	Time	\
0	Black	English	[Not Entered]	Los Angeles	1:00	
1	Hispanic/Latino	Spanish	[Not Entered]	Ventura	8:00	
2	White	English	[Not Entered]	Stanislaus	8:45	
3	Hispanic/Latino	English	[Not Entered]	Merced	9:00	
4	Hispanic/Latino	English	[Not Entered]	Stanislaus	9:00	
...	
52628	White	English	[Not Entered]	Kern	16:15	
52629	Hispanic/Latino	English	[Not Entered]	Tulare	16:30	
52630	Hispanic/Latino	English	[Not Entered]	Ventura	17:30	
52631	White	English	[Not Entered]	Stanislaus	17:30	
52632	Hispanic/Latino	English	[Not Entered]	Merced	9:00	

	date	month	dotw	int_dotw
0	2021-07-01	7	Thursday	3
1	2021-07-01	7	Thursday	3
2	2021-07-01	7	Thursday	3
3	2021-07-01	7	Thursday	3
4	2021-07-01	7	Thursday	3
...
52628	2024-06-28	6	Friday	4
52629	2024-06-28	6	Friday	4
52630	2024-06-28	6	Friday	4
52631	2024-06-28	6	Friday	4
52632	2024-06-30	6	Sunday	6

[52633 rows x 14 columns]

```
[25]: sns.countplot(data=cancelled, x='int_dotw', stat='percent')
plt.xticks([0, 1, 2, 3, 4, 5, 6], ['Monday', 'Tuesday', 'Wednesday',
    ↪ 'Thursday', 'Friday', 'Saturday', 'Sunday'])
plt.xticks(rotation=15)
plt.xlabel('Day of the Week');
```



```
[26]: # Looks like people tend to overall cancel in the middle of the week  
# We should look at getting rid of county disallowments the dataframe  
# take a look at the day of cancelations
```

```
[ ]:
```