

```
import pandas as pd

df = pd.read_csv('/content/diabetes.csv')

new_df = df.dropna()

print(new_df.to_string())
```

710	3	158	64.0	13	387	31.2	0.295	24	0
711	5	126	78.0	27	22	29.6	0.439	40	0
712	10	129	62.0	36	0	41.2	0.441	38	1
713	0	134	58.0	20	291	26.4	0.352	21	0
714	3	102	74.0	0	0	29.5	0.121	32	0
715	7	187	50.0	33	392	33.9	0.826	34	1
716	3	173	78.0	39	185	33.8	0.970	31	1
717	10	94	72.0	18	0	23.1	0.595	56	0
718	1	108	60.0	46	178	35.5	0.415	24	0
719	5	97	76.0	27	0	35.6	0.378	52	1
720	4	83	86.0	19	0	29.3	0.317	34	0
721	1	114	66.0	36	200	38.1	0.289	21	0
722	1	149	68.0	29	127	29.3	0.349	42	1
723	5	117	86.0	30	105	39.1	0.251	42	0
724	1	111	94.0	0	0	32.8	0.265	45	0
725	4	112	78.0	40	0	39.4	0.236	38	0
726	1	116	78.0	29	180	36.1	0.496	25	0
727	0	141	84.0	26	0	32.4	0.433	22	0
728	2	175	88.0	0	0	22.9	0.326	22	0
729	2	92	52.0	0	0	30.1	0.141	22	0
730	3	130	78.0	23	79	28.4	0.323	34	1
731	8	120	86.0	0	0	28.4	0.259	22	1
732	2	174	88.0	37	120	44.5	0.646	24	1
733	2	106	56.0	27	165	29.0	0.426	22	0
734	2	105	75.0	0	0	23.3	0.560	53	0
735	4	95	60.0	32	0	35.4	0.284	28	0
736	0	126	86.0	27	120	27.4	0.515	21	0
737	8	65	72.0	23	0	32.0	0.600	42	0
738	2	99	60.0	17	160	36.6	0.453	21	0
739	1	102	74.0	0	0	39.5	0.293	42	1
740	11	120	80.0	37	150	42.3	0.785	48	1
741	3	102	44.0	20	94	30.8	0.400	26	0
742	1	109	58.0	18	116	28.5	0.219	22	0
743	9	140	94.0	0	0	32.7	0.734	45	1
744	13	153	88.0	37	140	40.6	1.174	39	0
745	12	100	84.0	33	105	30.0	0.488	46	0
746	1	147	94.0	41	0	49.3	0.358	27	1
747	1	81	74.0	41	57	46.3	1.096	32	0
748	3	187	70.0	22	200	36.4	0.408	36	1
749	6	162	62.0	0	0	24.3	0.178	50	1
750	4	136	70.0	0	0	31.2	1.182	22	1
751	1	121	78.0	39	74	39.0	0.261	28	0
752	3	108	62.0	24	0	26.0	0.223	25	0
753	0	181	88.0	44	510	43.3	0.222	26	1
754	8	154	78.0	32	0	32.4	0.443	45	1
755	1	128	88.0	39	110	36.5	1.057	37	1
756	7	137	90.0	41	0	32.0	0.391	39	0
757	0	123	72.0	0	0	36.3	0.258	52	1
758	1	106	76.0	0	0	37.5	0.197	26	0
759	6	190	92.0	0	0	35.5	0.278	66	1
760	2	88	58.0	26	16	28.4	0.766	22	0
761	9	170	74.0	31	0	44.0	0.403	43	1
762	9	89	62.0	0	0	22.5	0.142	33	0
763	10	101	76.0	48	180	32.9	0.171	63	0
764	2	122	70.0	27	0	36.8	0.340	27	0
765	5	121	72.0	23	112	26.2	0.245	30	0
766	1	126	60.0	0	0	30.1	0.349	47	1
767	1	93	70.0	31	0	30.4	0.315	23	0

```
import pandas as pd

df = pd.read_csv('diabetes.csv')

df.dropna(inplace = True)

print(df.to_string())
```

8/23/23, 1:52 PM

DS\_exp\_01.ipynb - Colaboratory

722	1	149	68.0	29	127	29.3	0.349	42	1
723	5	117	86.0	30	105	39.1	0.251	42	0
724	1	111	94.0	0	0	32.8	0.265	45	0
725	4	112	78.0	40	0	39.4	0.236	38	0
726	1	116	78.0	29	180	36.1	0.496	25	0
727	0	141	84.0	26	0	32.4	0.433	22	0
728	2	175	88.0	0	0	22.9	0.326	22	0
729	2	92	52.0	0	0	30.1	0.141	22	0
730	3	130	78.0	23	79	28.4	0.323	34	1
731	8	120	86.0	0	0	28.4	0.259	22	1
732	2	174	88.0	37	120	44.5	0.646	24	1
733	2	106	56.0	27	165	29.0	0.426	22	0
734	2	105	75.0	0	0	23.3	0.560	53	0
735	4	95	60.0	32	0	35.4	0.284	28	0
736	0	126	86.0	27	120	27.4	0.515	21	0
737	8	65	72.0	23	0	32.0	0.600	42	0
738	2	99	60.0	17	160	36.6	0.453	21	0
739	1	102	74.0	0	0	39.5	0.293	42	1
740	11	120	80.0	37	150	42.3	0.785	48	1
741	3	102	44.0	20	94	30.8	0.400	26	0
742	1	109	58.0	18	116	28.5	0.219	22	0
743	9	140	94.0	0	0	32.7	0.734	45	1
744	13	153	88.0	37	140	40.6	1.174	39	0
745	12	100	84.0	33	105	30.0	0.488	46	0
746	1	147	94.0	41	0	49.3	0.358	27	1
747	1	81	74.0	41	57	46.3	1.096	32	0
748	3	187	70.0	22	200	36.4	0.408	36	1
749	6	162	62.0	0	0	24.3	0.178	50	1
750	4	136	70.0	0	0	31.2	1.182	22	1
751	1	121	78.0	39	74	39.0	0.261	28	0
752	3	108	62.0	24	0	26.0	0.223	25	0
753	0	181	88.0	44	510	43.3	0.222	26	1
754	8	154	78.0	32	0	32.4	0.443	45	1
755	1	128	88.0	39	110	36.5	1.057	37	1
756	7	137	90.0	41	0	32.0	0.391	39	0
757	0	123	72.0	0	0	36.3	0.258	52	1
758	1	106	76.0	0	0	37.5	0.197	26	0
759	6	190	92.0	0	0	35.5	0.278	66	1
760	2	88	58.0	26	16	28.4	0.766	22	0
761	9	170	74.0	31	0	44.0	0.403	43	1
762	9	89	62.0	0	0	22.5	0.142	33	0
763	10	101	76.0	48	180	32.9	0.171	63	0
764	2	122	70.0	27	0	36.8	0.340	27	0
765	5	121	72.0	23	112	26.2	0.245	30	0
766	1	126	60.0	0	0	30.1	0.349	47	1
767	1	93	70.0	31	0	30.4	0.315	23	0

```
df.fillna(130, inplace = True)
print(df.to_string())
```

747	1	81	74.0	41	37	40.3	1.057	37	1
748	3	187	70.0	22	200	36.4	0.408	36	1
749	6	162	62.0	0	0	24.3	0.178	50	1
750	4	136	70.0	0	0	31.2	1.182	22	1
751	1	121	78.0	39	74	39.0	0.261	28	0
752	3	108	62.0	24	0	26.0	0.223	25	0
753	0	181	88.0	44	510	43.3	0.222	26	1
754	8	154	78.0	32	0	32.4	0.443	45	1
755	1	128	88.0	39	110	36.5	1.057	37	1
756	7	137	90.0	41	0	32.0	0.391	39	0
757	0	123	72.0	0	0	36.3	0.258	52	1
758	1	106	76.0	0	0	37.5	0.197	26	0
759	6	190	92.0	0	0	35.5	0.278	66	1
760	2	88	58.0	26	16	28.4	0.766	22	0
761	9	170	74.0	31	0	44.0	0.403	43	1
762	9	89	62.0	0	0	22.5	0.142	33	0
763	10	101	76.0	48	180	32.9	0.171	63	0
764	2	122	70.0	27	0	36.8	0.340	27	0
765	5	121	72.0	23	112	26.2	0.245	30	0
766	1	126	60.0	0	0	30.1	0.349	47	1
767	1	93	70.0	31	0	30.4	0.315	23	0

```
import pandas as pd
```

```
df = pd.read_csv('diabetes.csv')
```

```
df["Glucose"].fillna(130, inplace = True)
print(df.to_string())
```

710	3	158	64.0	13	387	31.2	0.295	24	0
711	5	126	78.0	27	22	29.6	0.439	40	0
712	10	129	62.0	36	0	41.2	0.441	38	1
713	0	134	58.0	20	291	26.4	0.352	21	0
714	3	102	74.0	0	0	29.5	0.121	32	0
715	7	187	50.0	33	392	33.9	0.826	34	1
716	3	173	78.0	39	185	33.8	0.970	31	1
717	10	94	72.0	18	0	23.1	0.595	56	0
718	1	108	60.0	46	178	35.5	0.415	24	0
719	5	97	76.0	27	0	35.6	0.378	52	1
720	4	83	86.0	19	0	29.3	0.317	34	0
721	1	114	66.0	36	200	38.1	0.289	21	0
722	1	149	68.0	29	127	29.3	0.349	42	1
723	5	117	86.0	30	105	39.1	0.251	42	0
724	1	111	94.0	0	0	32.8	0.265	45	0
725	4	112	78.0	40	0	39.4	0.236	38	0
726	1	116	78.0	29	180	36.1	0.496	25	0
727	0	141	84.0	26	0	32.4	0.433	22	0
728	2	175	88.0	0	0	22.9	0.326	22	0
729	2	92	52.0	0	0	30.1	0.141	22	0
730	3	130	78.0	23	79	28.4	0.323	34	1
731	8	120	86.0	0	0	28.4	0.259	22	1
732	2	174	88.0	37	120	44.5	0.646	24	1
733	2	106	56.0	27	165	29.0	0.426	22	0
734	2	105	75.0	0	0	23.3	0.560	53	0
735	4	95	60.0	32	0	35.4	0.284	28	0
736	0	126	86.0	27	120	27.4	0.515	21	0
737	8	65	72.0	23	0	32.0	0.600	42	0
738	2	99	60.0	17	160	36.6	0.453	21	0
739	1	102	74.0	0	0	39.5	0.293	42	1
740	11	120	80.0	37	150	42.3	0.785	48	1
741	3	102	44.0	20	94	30.8	0.400	26	0
742	1	109	58.0	18	116	28.5	0.219	22	0
743	9	140	94.0	0	0	32.7	0.734	45	1
744	13	153	88.0	37	140	40.6	1.174	39	0
745	12	100	84.0	33	105	30.0	0.488	46	0
746	1	147	94.0	41	0	49.3	0.358	27	1
747	1	81	74.0	41	57	46.3	1.096	32	0
748	3	187	70.0	22	200	36.4	0.408	36	1
749	6	162	62.0	0	0	24.3	0.178	50	1
750	4	136	70.0	0	0	31.2	1.182	22	1
751	1	121	78.0	39	74	39.0	0.261	28	0
752	3	108	62.0	24	0	26.0	0.223	25	0
753	0	181	88.0	44	510	43.3	0.222	26	1
754	8	154	78.0	32	0	32.4	0.443	45	1
755	1	128	88.0	39	110	36.5	1.057	37	1
756	7	137	90.0	41	0	32.0	0.391	39	0
757	0	123	72.0	0	0	36.3	0.258	52	1
758	1	106	76.0	0	0	37.5	0.197	26	0
759	6	190	92.0	0	0	35.5	0.278	66	1
760	2	88	58.0	26	16	28.4	0.766	22	0
761	9	170	74.0	31	0	44.0	0.403	43	1
762	9	89	62.0	0	0	22.5	0.142	33	0
763	10	101	76.0	48	180	32.9	0.171	63	0
764	2	122	70.0	27	0	36.8	0.340	27	0
765	5	121	72.0	23	112	26.2	0.245	30	0
766	1	126	60.0	0	0	30.1	0.349	47	1
767	1	93	70.0	31	0	30.4	0.315	23	0

```
x = df["BloodPressure"].mean()
```

```
df["BloodPressure"].fillna(x, inplace = True)
print(df.to_string())
```

710	3	158	64.000000	13	387	31.2	0.295	24	0
711	5	126	78.000000	27	22	29.6	0.439	40	0
712	10	129	62.000000	36	0	41.2	0.441	38	1
713	0	134	58.000000	20	291	26.4	0.352	21	0
714	3	102	74.000000	0	0	29.5	0.121	32	0
715	7	187	50.000000	33	392	33.9	0.826	34	1
716	3	173	78.000000	39	185	33.8	0.970	31	1
717	10	94	72.000000	18	0	23.1	0.595	56	0
718	1	108	60.000000	46	178	35.5	0.415	24	0
719	5	97	76.000000	27	0	35.6	0.378	52	1
720	4	83	86.000000	19	0	29.3	0.317	34	0
721	1	114	66.000000	36	200	38.1	0.289	21	0
722	1	149	68.000000	29	127	29.3	0.349	42	1
723	5	117	86.000000	30	105	39.1	0.251	42	0
724	1	111	94.000000	0	0	32.8	0.265	45	0
725	4	112	78.000000	40	0	39.4	0.236	38	0
726	1	116	78.000000	29	180	36.1	0.496	25	0
727	0	141	84.000000	26	0	32.4	0.433	22	0
728	2	175	88.000000	0	0	22.9	0.326	22	0
729	2	92	52.000000	0	0	30.1	0.141	22	0
730	3	130	78.000000	23	79	28.4	0.323	34	1
731	8	120	86.000000	0	0	28.4	0.259	22	1
732	2	174	88.000000	37	120	44.5	0.646	24	1
733	2	106	56.000000	27	165	29.0	0.426	22	0
734	2	105	75.000000	0	0	23.3	0.560	53	0
735	4	95	60.000000	32	0	35.4	0.284	28	0
736	0	126	86.000000	27	120	27.4	0.515	21	0
737	8	65	72.000000	23	0	32.0	0.600	42	0
738	2	99	60.000000	17	160	36.6	0.453	21	0
739	1	102	74.000000	0	0	39.5	0.293	42	1
740	11	120	80.000000	37	150	42.3	0.785	48	1
741	3	102	44.000000	20	94	30.8	0.400	26	0
742	1	109	58.000000	18	116	28.5	0.219	22	0
743	9	140	94.000000	0	0	32.7	0.734	45	1
744	13	153	88.000000	37	140	40.6	1.174	39	0
745	12	100	84.000000	33	105	30.0	0.488	46	0
746	1	147	94.000000	41	0	49.3	0.358	27	1
747	1	81	74.000000	41	57	46.3	1.096	32	0
748	3	187	70.000000	22	200	36.4	0.408	36	1
749	6	162	62.000000	0	0	24.3	0.178	50	1
750	4	136	70.000000	0	0	31.2	1.182	22	1
751	1	121	78.000000	39	74	39.0	0.261	28	0
752	3	108	62.000000	24	0	26.0	0.223	25	0
753	0	181	88.000000	44	510	43.3	0.222	26	1
754	8	154	78.000000	32	0	32.4	0.443	45	1
755	1	128	88.000000	39	110	36.5	1.057	37	1
756	7	137	90.000000	41	0	32.0	0.391	39	0
757	0	123	72.000000	0	0	36.3	0.258	52	1
758	1	106	76.000000	0	0	37.5	0.197	26	0
759	6	190	92.000000	0	0	35.5	0.278	66	1
760	2	88	58.000000	26	16	28.4	0.766	22	0
761	9	170	74.000000	31	0	44.0	0.403	43	1
762	9	89	62.000000	0	0	22.5	0.142	33	0
763	10	101	76.000000	48	180	32.9	0.171	63	0
764	2	122	70.000000	27	0	36.8	0.340	27	0
765	5	121	72.000000	23	112	26.2	0.245	30	0
766	1	126	60.000000	0	0	30.1	0.349	47	1
767	1	93	70.000000	31	0	30.4	0.315	23	0

```
x = df["BloodPressure"].median()
```

```
df["BloodPressure"].fillna(x, inplace = True)
print(df.to_string())
```

69	4	146	85.000000	27	100	28.9	0.189	27	0
70	2	100	66.000000	20	90	32.9	0.867	28	1
71	5	139	64.000000	35	140	28.6	0.411	26	0
72	13	126	90.000000	0	0	43.4	0.583	42	1
73	4	129	86.000000	20	270	35.1	0.231	23	0
74	1	79	75.000000	30	0	32.0	0.396	22	0
75	1	0	48.000000	20	0	24.7	0.140	22	0
76	7	62	78.000000	0	0	32.6	0.391	41	0
77	5	95	72.000000	33	0	37.7	0.370	27	0
78	0	131	69.549202	0	0	43.2	0.270	26	1
79	2	112	66.000000	22	0	25.0	0.307	24	0
80	3	113	44.000000	13	0	22.4	0.140	22	0
81	2	74	0.000000	0	0	0.0	0.102	22	0
82	7	83	78.000000	26	71	29.3	0.767	36	0
83	0	101	65.000000	28	0	24.6	0.237	22	0
84	5	137	108.000000	0	0	48.8	0.227	37	1
85	2	110	74.000000	29	125	32.4	0.698	27	0
86	13	106	72.000000	54	0	36.6	0.178	45	0
87	2	100	68.000000	25	71	38.5	0.324	26	0
88	15	136	70.000000	32	110	37.1	0.153	43	1
89	1	107	68.000000	19	0	26.5	0.165	24	0
90	1	80	55.000000	0	0	19.1	0.258	21	0
91	4	123	80.000000	15	176	32.0	0.443	34	0
92	7	81	78.000000	40	48	46.7	0.261	42	0
93	4	134	72.000000	0	0	23.8	0.277	60	1
94	2	142	82.000000	18	64	24.7	0.761	21	0
95	6	144	72.000000	27	228	33.9	0.255	40	0
96	2	92	62.000000	28	0	31.6	0.130	24	0
97	1	71	48.000000	18	76	20.4	0.323	22	0
98	6	93	50.000000	30	64	28.7	0.356	23	0
99	1	122	90.000000	51	220	49.7	0.325	31	1
100	1	163	72.000000	0	0	39.0	1.222	33	1
101	1	151	60.000000	0	0	26.1	0.179	22	0
102	0	125	96.000000	0	0	22.5	0.262	21	0
103	1	81	72.000000	18	40	26.6	0.283	24	0
104	2	85	65.000000	0	0	39.6	0.930	27	0
105	1	126	56.000000	29	152	28.7	0.801	21	0
106	1	96	122.000000	0	0	22.4	0.207	27	0
107	4	144	58.000000	28	140	29.5	0.287	37	0
108	3	83	58.000000	31	18	34.3	0.336	25	0
109	0	95	85.000000	25	36	37.4	0.247	24	1
110	0	171	70.000000	22	135	22.2	0.100	24	1

```
import pandas as pd

df = pd.read_csv('diabetes.csv')

x = df["BloodPressure"].mode()

df["BloodPressure"].fillna(x, inplace = True)

df.loc[7, 'Duration'] = 45
print(df.to_string())
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome	Duration
0	6	148	72.0	35	0	33.6	0.627	50	1	NaN
1	1	85	66.0	29	0	26.6	0.351	31	0	NaN
2	8	183	64.0	0	0	23.3	0.672	32	1	NaN
3	1	89	66.0	23	94	28.1	0.167	21	0	NaN
4	0	137	40.0	35	168	43.1	2.288	33	1	NaN
5	5	116	74.0	0	0	25.6	0.201	30	0	NaN
6	3	78	50.0	32	88	31.0	0.248	26	1	NaN
7	10	115	NaN	0	0	35.3	0.134	29	0	45.0
8	2	197	70.0	45	543	30.5	0.158	53	1	NaN
9	8	125	96.0	0	0	0.0	0.232	54	1	NaN
10	4	110	92.0	0	0	37.6	0.191	30	0	NaN
11	10	168	74.0	0	0	38.0	0.537	34	1	NaN
12	10	139	80.0	0	0	27.1	1.441	57	0	NaN
13	1	189	60.0	23	846	30.1	0.398	59	1	NaN
14	5	166	72.0	19	175	25.8	0.587	51	1	NaN
15	7	100	NaN	0	0	30.0	0.484	32	1	NaN
16	0	118	84.0	47	230	45.8	0.551	31	1	NaN
17	7	107	74.0	0	0	29.6	0.254	31	1	NaN
18	1	103	30.0	38	83	43.3	0.183	33	0	NaN
19	1	115	70.0	30	96	34.6	0.529	32	1	NaN
20	3	126	88.0	41	235	39.3	0.704	27	0	NaN
21	8	99	NaN	0	0	35.4	0.388	50	0	NaN
22	7	196	90.0	0	0	39.8	0.451	41	1	NaN
23	9	119	80.0	35	0	29.0	0.263	29	1	NaN
24	11	143	94.0	33	146	36.6	0.254	51	1	NaN
25	10	125	70.0	26	115	31.1	0.205	41	1	NaN
26	7	147	76.0	0	0	39.4	0.257	43	1	NaN
27	1	97	66.0	15	140	23.2	0.487	22	0	NaN
28	13	145	82.0	19	110	22.2	0.245	57	0	NaN
29	5	117	92.0	0	0	34.1	0.337	38	0	NaN
30	5	109	75.0	26	0	36.0	0.546	60	0	NaN
31	3	158	76.0	36	245	31.6	0.851	28	1	NaN
32	3	88	58.0	11	54	24.8	0.267	22	0	NaN
33	6	92	92.0	0	0	19.9	0.188	28	0	NaN

34	10	122	78.0	31	0	27.6	0.512	45	0	NaN
35	4	103	60.0	33	192	24.0	0.966	33	0	NaN
36	11	138	76.0	0	0	33.2	0.420	35	0	NaN
37	9	102	NaN	37	0	32.9	0.665	46	1	NaN
38	2	90	68.0	42	0	38.2	0.503	27	1	NaN
39	4	111	72.0	47	207	37.1	1.390	56	1	NaN
40	3	180	64.0	25	70	34.0	0.271	26	0	NaN
41	7	133	84.0	0	0	40.2	0.696	37	0	NaN
42	7	106	92.0	18	0	22.7	0.235	48	0	NaN
43	9	171	110.0	24	240	45.4	0.721	54	1	NaN
44	7	159	64.0	0	0	27.4	0.294	40	0	NaN
45	0	180	66.0	39	0	42.0	1.893	25	1	NaN
46	1	146	56.0	0	0	29.7	0.564	29	0	NaN
47	2	71	70.0	27	0	28.0	0.586	22	0	NaN
48	7	103	66.0	32	0	39.1	0.344	31	1	NaN
49	7	105	0.0	0	0	0.0	0.305	24	0	NaN
50	1	103	80.0	11	82	19.4	0.491	22	0	NaN
51	1	101	50.0	15	36	24.2	0.526	26	0	NaN
52	5	88	66.0	21	23	24.4	0.342	30	0	NaN
53	8	176	90.0	34	300	33.7	0.467	58	1	NaN
54	7	150	66.0	42	342	34.7	0.718	42	0	NaN
55	1	73	50.0	10	0	23.0	0.248	21	0	NaN
56	7	187	68.0	39	304	37.7	0.254	41	1	NaN

```
for x in df.index:
    if df.loc[x, "Duration"] > 120:
        df.loc[x, "Duration"] = 120
```

```
for x in df.index:
    if df.loc[x, "Duration"] > 120:
        df.drop(x, inplace = True)
print(df.to_string())
```

```
print(df.duplicated())
```

0	False
1	False
2	False
3	False
4	False
...	
763	False
764	False
765	False
766	False
767	False
Length: 768, dtype: bool	

```
df.drop_duplicates(inplace = True)
```

```
print(df.to_string())
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome	Duration
0	6	148	72.0	35	0	33.6	0.627	50	1	NaN
1	1	85	66.0	29	0	26.6	0.351	31	0	NaN
2	8	183	64.0	0	0	23.3	0.672	32	1	NaN
3	1	89	66.0	23	94	28.1	0.167	21	0	NaN
4	0	137	40.0	35	168	43.1	2.288	33	1	NaN
5	5	116	74.0	0	0	25.6	0.201	30	0	NaN
6	3	78	50.0	32	88	31.0	0.248	26	1	NaN
7	10	115	NaN	0	0	35.3	0.134	29	0	45.0
8	2	197	70.0	45	543	30.5	0.158	53	1	NaN
9	8	125	96.0	0	0	0.0	0.232	54	1	NaN
10	4	110	92.0	0	0	37.6	0.191	30	0	NaN
11	10	168	74.0	0	0	38.0	0.537	34	1	NaN
12	10	139	80.0	0	0	27.1	1.441	57	0	NaN
13	1	189	60.0	23	846	30.1	0.398	59	1	NaN
14	5	166	72.0	19	175	25.8	0.587	51	1	NaN
15	7	100	NaN	0	0	30.0	0.484	32	1	NaN
16	0	118	84.0	47	230	45.8	0.551	31	1	NaN
17	7	107	74.0	0	0	29.6	0.254	31	1	NaN
18	1	103	30.0	38	83	43.3	0.183	33	0	NaN
19	1	115	70.0	30	96	34.6	0.529	32	1	NaN
20	3	126	88.0	41	235	39.3	0.704	27	0	NaN
21	8	99	NaN	0	0	35.4	0.388	50	0	NaN
22	7	196	90.0	0	0	39.8	0.451	41	1	NaN
23	9	119	80.0	35	0	29.0	0.263	29	1	NaN
24	11	143	94.0	33	146	36.6	0.254	51	1	NaN
25	10	125	70.0	26	115	31.1	0.205	41	1	NaN
26	7	147	76.0	0	0	39.4	0.257	43	1	NaN
27	1	97	66.0	15	140	23.2	0.487	22	0	NaN
28	13	145	82.0	19	110	22.2	0.245	57	0	NaN
29	5	117	92.0	0	0	34.1	0.337	38	0	NaN
30	5	109	75.0	26	0	36.0	0.546	60	0	NaN
31	3	158	76.0	36	245	31.6	0.851	28	1	NaN
32	3	88	58.0	11	54	24.8	0.267	22	0	NaN

8/23/23, 1:52 PM

DS\_exp\_01.ipynb - Colaboratory

33	6	92	92.0	0	0	19.9	0.188	28	0	NaN
34	10	122	78.0	31	0	27.6	0.512	45	0	NaN
35	4	103	60.0	33	192	24.0	0.966	33	0	NaN
36	11	138	76.0	0	0	33.2	0.420	35	0	NaN
37	9	102	NaN	37	0	32.9	0.665	46	1	NaN
38	2	90	68.0	42	0	38.2	0.503	27	1	NaN
39	4	111	72.0	47	207	37.1	1.390	56	1	NaN
40	3	180	64.0	25	70	34.0	0.271	26	0	NaN
41	7	133	84.0	0	0	40.2	0.696	37	0	NaN
42	7	106	92.0	18	0	22.7	0.235	48	0	NaN
43	9	171	110.0	24	240	45.4	0.721	54	1	NaN
44	7	159	64.0	0	0	27.4	0.294	40	0	NaN
45	0	180	66.0	39	0	42.0	1.893	25	1	NaN
46	1	146	56.0	0	0	29.7	0.564	29	0	NaN
47	2	71	70.0	27	0	28.0	0.586	22	0	NaN
48	7	103	66.0	32	0	39.1	0.344	31	1	NaN
49	7	105	0.0	0	0	0.0	0.305	24	0	NaN
50	1	103	80.0	11	82	19.4	0.491	22	0	NaN
51	1	101	50.0	15	36	24.2	0.526	26	0	NaN
52	5	88	66.0	21	23	24.4	0.342	30	0	NaN
53	8	176	90.0	34	300	33.7	0.467	58	1	NaN
54	7	150	66.0	42	342	34.7	0.718	42	0	NaN
55	1	73	50.0	10	0	23.0	0.248	21	0	NaN
56	7	187	68.0	39	304	37.7	0.254	41	1	NaN

```
# drop the "Cabin" column from the dataframe
titanic_data = titanic_data.drop(columns='Duration', axis=1)
```