

IE 5390 – Assignment - 01

Name: Simran Abhay Sinha

Lab 01

2. Clean the data so that all the attributes have their own columns, i.e., 12 columns.

Clean the header row so that only the attribute names are displayed, without any quotation marks or special characters.

Result:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality		
2	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6		
3	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6		
4	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6		
5	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6		
6	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6		
7	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6		
8	6.2	0.32	0.16	7	0.045	30	136	0.9949	3.18	0.47	9.6	6		
9	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6		
10	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6		
11	8.1	0.22	0.43	1.5	0.044	28	129	0.9938	3.22	0.45	11	6		
12	8.1	0.27	0.41	1.45	0.033	11	63	0.9908	3.26	0.44	10.1	6		
13	8.6	0.23	0.4	4.2	0.035	17	109	0.9947	3.14	0.53	9.7	5		
14	7.9	0.18	0.37	1.2	0.04	16	75	0.992	3.18	0.63	10.8	5		
15	6.6	0.16	0.4	1.5	0.044	48	143	0.9912	3.54	0.52	12.4	7		
16	8.3	0.42	0.62	19.25	0.04	41	172	1.0002	2.98	0.67	9.7	5		
17	6.6	0.17	0.38	1.5	0.042	20	123	0.9914	3.45	0.55	11.0	7		
18	6.3	0.48	0.04	1.1	0.046	30	99	0.9928	3.24	0.36	9.6	6		
19	White	6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8	
20	White	7.4	0.24	0.43	1.1	0.033	17	171	0.9917	3.31	0.53	11.6	6	
21	White	6.5	0.31	0.14	7.5	0.044	34	133	0.9965	3.22	0.5	9.5	5	
22	White	6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8	
23	White	6.4	0.31	0.39	2.9	0.048	19	102	0.9912	3.31	0.35	11.7	7	
24	White	6.8	0.26	0.42	1.7	0.049	41	122	0.993	3.47	0.48	10.5	8	
25	White	6.6	0.67	0.14	1.5	0.074	25	108	0.9937	3.05	0.51	9.3	5	
26	White	6.9	0.27	0.41	1.3	0.042	18	142	0.9903	3.44	0.47	10.6	6	
27	White	7	0.25	0.32	9	0.046	56	245	0.9965	3.25	0.5	10.4	6	
28	White	6.5	0.54	0.35	1.2	0.052	35	146	0.9945	3.45	0.44	10	6	
29	White	7	0.29	0.39	8.7	0.051	35	141	0.9961	3.48	0.53	10	6	
30	White	7.4	0.27	0.48	1.1	0.047	17	132	0.9914	3.19	0.49	11.6	6	
31	White	7.2	0.32	0.36	1.2	0.053	27	114	0.9946	3.1	0.71	12.2	7	
32	White	8.5	0.24	0.39	10.4	0.044	20	142	0.9974	3.2	0.53	10	6	
33	White	8.3	0.14	0.34	1.1	0.042	7	47	0.9834	3.47	0.4	10.2	6	
34	White	7.4	0.26	0.36	2.05	0.05	31	100	0.9952	3.44	0.44	10.6	6	
35	White	6.2	0.12	0.34	1.5	0.045	43	117	0.9939	3.42	0.51	9	6	
36	White	5.8	0.27	0.24	14.95	0.044	22	179	0.9962	3.37	0.37	10.2	6	
37	White	7.3	0.29	0.43	1.7	0.048	21	123	0.9905	3.41	0.42	12.5	5	
38	White	6.5	0.39	0.23	5.4	0.051	25	149	0.9934	3.24	0.35	10	5	
39	White	7	0.33	0.32	1.2	0.053	38	138	0.9906	3.13	0.28	10.4	6	
40	White	7.3	0.24	0.39	17.55	0.047	40	140	0.9909	3.21	0.36	8.6	5	
41	White	7.4	0.24	0.39	17.55	0.047	41	129	0.9909	3.21	0.36	8.6	5	

3. Create a column called “Wine_Type” and place it right at the beginning of both red and white wine files. For the red wine data, add the label “Red” under “Wine_Type”. Similarly, for the white wine data, add the label “White” under “Wine_Type”. Both datasets should now consist of 13 columns.

Result:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Wine_type	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality							
2	White	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6							
3	White	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6							
4	White	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6							
5	White	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.44	9.9	6							
6	White	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.44	9.9	6							
7	White	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6							
8	White	6.2	0.32	0.16	7	0.045	30	136	0.9949	3.17	0.47	9.6	6							
9	White	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6							
10	White	6.3	0.3	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6								
11	White	8.1	0.22	0.43	1.5	0.044	28	109	0.9908	3.45	0.45	11	6							
12	White	8.1	0.27	0.41	1.45	0.033	11	63	0.9908	2.99	0.56	12	5							
13	White	8.6	0.23	0.4	4.2	0.035	16	109	0.9947	3.14	0.53	9.7	5							
14	White	7.9	0.18	0.37	1.2	0.04	16	75	0.992	3.18	0.63	10.8	5							
15	White	6.6	0.16	0.4	1.5	0.044	48	143	0.9912	3.54	0.52	12.4	7							
16	White	8.3	0.42	0.62	19.25	0.04	41	172	1.0002	2.98	0.67	9.7	5							
17	White	6.6	0.17	0.38	1.5	0.042	20	123	0.9914	3.45	0.55	11.0	7							
18	White	6.3	0.48	0.04	1.1	0.046	30	99	0.9928	3.24	0.36	9.6	6							
19	White	6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8							
20	White	7.4	0.24	0.43	1.1	0.033	17	171	0.9917	3.31	0.53	11.6	6							
21	White	6.5	0.31	0.14	7.5	0.044	34	133	0.9965	3.22	0.5	9.5	5							
22	White	6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8							
23	White	6.4	0.31	0.39	2.9	0.048	19	102	0.9912	3.31	0.35	11.7	7							
24	White	6.8	0.26	0.42	1.7	0.049	41	122	0.993	3.47	0.48	10.5	8							
25	White	6.6	0.67	0.14	1.5	0.074	25	108	0.9937	3.05	0.51	9.3	5							
26	White	6.9	0.27	0.41	1.3	0.042	18	142	0.9903	3.44	0.47	10.6	6							
27	White	7	0.25	0.32	9	0.046	56	245	0.9965	3.25	0.5	10.4	6							
28	White	6.5	0.54	0.35	1.2	0.052	35	146	0.9945	3.45	0.44	10	6							
29	White	7	0.29	0.39	8.7	0.051	35	141	0.9961	3.48	0.53	10	6							
30	White	7.4	0.27	0.48	1.1	0.047	17	132	0.9914	3.19	0.49	11.6	6							
31	White	7.2	0.32	0.36	1.2	0.053	27	114	0.9946	3.1	0.71	12.2	7							
32	White	8.5	0.24	0.39	10.4	0.044	20	142	0.9974	3.2	0.53	10	6							
33	White	8.3	0.14	0.34	1.1	0.042	7	47	0.9834	3.47	0.4	10.2	6							
34	White	7.4	0.26	0.36	2.05	0.05	31	100	0.9952	3.44	0.44	10.6	6							
35	White	6.2	0.12	0.34	1.5	0.045	43	117	0.											

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Wine_type	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality		
2	Red	7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5		
3	Red	7.8	0.88	0	2.6	0.098	25	67	0.9968	3.2	0.68	9.8	5		
4	Red	7.8	0.76	0.04	2.3	0.092	15	54	0.997	3.26	0.65	9.8	5		
5	Red	11.2	0.28	0.56	1.9	0.075	17	60	0.998	3.16	0.58	9.8	6		
6	Red	7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5		
7	Red	7.4	0.66	0	1.8	0.075	13	40	0.9978	3.51	0.56	9.4	5		
8	Red	7.9	0.6	0.06	1.6	0.069	15	59	0.9964	3.3	0.46	9.4	5		
9	Red	7.3	0.65	0	1.2	0.065	15	21	0.9946	3.39	0.47	10	7		
10	Red	7.8	0.58	0.02	2	0.073	9	18	0.9968	3.36	0.57	9.5	7		
11	Red	7.5	0.5	0.36	6.1	0.071	17	102	0.9978	3.35	0.8	10.5	5		
12	Red	6.7	0.58	0.08	1.8	0.097	15	65	0.9959	3.28	0.54	9.2	5		
13	Red	7.5	0.5	0.36	6.1	0.071	17	102	0.9978	3.35	0.8	10.5	5		
14	Red	5.6	0.615	0	1.6	0.089	16	59	0.9943	3.58	0.52	9.9	5		
15	Red	7.8	0.61	0.29	1.6	0.114	9	29	0.9974	3.26	1.56	9.1	5		
16	Red	8.9	0.62	0.18	3.8	0.176	52	145	0.9986	3.16	0.88	9.2	5		
17	Red	8.9	0.62	0.19	3.9	0.17	51	148	0.9986	3.17	0.93	9.2	5		
18	Red	8.5	0.28	0.56	1.8	0.092	35	103	0.9969	3.3	0.75	10.5	7		
19	Red	8.1	0.56	0.28	1.7	0.368	16	56	0.9968	3.11	1.28	9.3	5		
20	Red	7.4	0.59	0.08	4.4	0.086	6	29	0.9974	3.38	0.5	9	4		
21	Red	7.9	0.32	0.51	1.8	0.341	17	56	0.9969	3.04	1.08	9.2	6		
22	Red	8.9	0.22	0.48	1.8	0.077	29	60	0.9968	3.39	0.53	9.4	6		
23	Red	7.6	0.39	0.31	2.3	0.082	23	71	0.9982	3.52	0.65	9.7	5		
24	Red	7.9	0.43	0.21	1.6	0.106	10	37	0.9966	3.17	0.91	9.5	5		
25	Red	8.5	0.49	0.11	2.3	0.084	9	67	0.9968	3.17	0.53	9.4	5		
26	Red	6.9	0.4	0.14	2.4	0.085	21	40	0.9968	3.43	0.63	9.7	6		
27	Red	6.3	0.39	0.16	1.4	0.08	11	23	0.9955	3.34	0.56	9.3	5		
28	Red	7.6	0.41	0.24	1.8	0.08	4	11	0.9962	3.28	0.59	9.5	5		
29	Red	7.9	0.43	0.21	1.6	0.106	10	37	0.9966	3.17	0.91	9.5	5		
30	Red	7.1	0.71	0	1.9	0.08	14	35	0.9972	3.47	0.55	9.4	5		

4. Combine both the red and white wine datasets. Combined dataset should have all red wine records first and then the white wine records. There will be a total of 6,497 records excluding the header row. Name the dataset “wine_combined.xlsx”

Result:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Wine_type	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality							
2	White	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6							
3	White	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6							
4	White	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6							
5	White	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6							
6	White	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6							
7	White	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6							
8	White	6.2	0.32	0.16	7	0.045	30	136	0.9949	3.18	0.47	9.6	6							
9	White	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6							
10	White	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6							
11	White	8.1	0.22	0.43	1.5	0.044	28	129	0.9938	3.22	0.45	11	6							
12	White	8.1	0.27	0.41	1.45	0.033	11	63	0.9908	2.99	0.56	12	5							
13	White	8.6	0.23	0.4	4.2	0.035	17	109	0.9947	3.14	0.53	9.7	5							
14	White	7.9	0.18	0.37	1.2	0.04	16	75	0.992	3.18	0.63	10.8	5							
15	White	6.6	0.16	0.4	1.5	0.044	48	143	0.9912	3.54	0.52	12.4	7							
16	White	8.3	0.42	0.62	19.25	0.04	41	172	1.0002	2.98	0.67	9.7	5							
17	White	6.5	0.17	0.38	1.5	0.02	28	112	0.9944	3.16	0.55	11.4	7							
18	White	6.3	0.48	0.04	1.1	0.046	36	99	0.9929	3.24	0.56	9.8	6							
19	White	6.2	0.69	0.49	1.2	0.029	29	75	0.9992	3.33	0.39	12.8	8							
20	White	7.4	0.34	0.42	1.1	0.033	17	171	0.9917	3.12	0.53	11.3	6							
21	White	6.5	0.31	0.14	7.5	0.044	34	133	0.9965	3.22	0.5	9.5	5							
22	White	6.2	0.66	0.49	1.2	0.029	29	75	0.9982	3.33	0.39	12.8	8							
23	White	6.4	0.31	0.38	2.9	0.038	19	102	0.9912	3.17	0.35	11	7							
24	White	6.8	0.26	0.42	1.7	0.049	41	122	0.993	3.47	0.48	10.5	8							
25	White	7.6	0.67	0.14	1.5	0.074	25	168	0.9937	3.05	0.51	9.3	5							
26	White	6.6	0.27	0.41	1.3	0.052	16	142	0.9951	3.42	0.47	10	6							
27	White	7	0.25	0.32	9	0.046	56	245	0.9955	3.25	0.5	10.4	6							
28	White	6.9	0.24	0.35	1	0.052	35	146	0.993	3.45	0.44	10	6							
29	White	7	0.28	0.39	8.7	0.051	32	141	0.9961	3.38	0.53	10.5	6							
30	White	7.4	0.77	0.48	1.1	0.047	17	132	0.9914	3.19	0.49	11.6	8							

5. Check the data for missing values using both the methods discussed in class

Result:

6. Remove any leading and trailing spaces that the data and headers may have

Result:

winequality-white														
Home	Insert	Draw	Page Layout	Formulas	Data	Review	View	Automate	Arial	14	A ⁺	A ⁻	General	
Paste	Cut	Copy	Format Cells	Conditional Formatting	Format as Table	Cell Styles	Insert	Delete	Format	Sort & Filter	Find & Select	Sensi		
A1									=TRIM(wine_combined!A1:M1)					
1	Wine type	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	
2	White	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6	
3	White	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6	
4	White	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6	
5	White	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6	
6	White	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6	
7	White	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6	
8	White	6.2	0.32	0.16	7	0.045	30	136	0.9949	3.18	0.47	9.6	6	
9	White	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6	
10	White	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6	
11	White	8.1	0.22	0.43	1.5	0.044	28	129	0.9938	3.22	0.45	11	6	
12	White	8.1	0.27	0.41	1.45	0.033	11	63	0.9908	2.99	0.56	12	5	
13	White	8.6	0.23	0.4	4.2	0.035	17	109	0.9947	3.14	0.53	9.7	5	
14	White	7.9	0.18	0.37	1.2	0.04	16	75	0.992	3.18	0.63	10.8	5	
15	White	6.6	0.16	0.4	1.5	0.044	48	143	0.9912	3.54	0.52	12.4	7	
16	White	8.3	0.42	0.62	19.25	0.04	41	172	1.0002	2.98	0.67	9.7	5	
17	White	6.6	0.17	0.38	1.5	0.032	28	112	0.9914	3.25	0.55	11.4	7	
18	White	6.3	0.48	0.04	1.1	0.046	30	99	0.9928	3.24	0.36	9.6	6	
19	White	6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8	
20	White	7.4	0.34	0.42	1.1	0.033	17	171	0.9917	3.12	0.53	11.3	6	
21	White	6.5	0.31	0.14	7.5	0.044	34	133	0.9955	3.22	0.5	9.5	5	
22	White	6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8	
23	White	6.4	0.31	0.38	2.9	0.038	19	102	0.9912	3.17	0.35	11	7	
24	White	6.8	0.26	0.42	1.7	0.049	41	122	0.993	3.47	0.48	10.5	8	

IE5390 Structured Data Analytics

Screenshot of Microsoft Excel showing the formula bar with the formula `=VALUE(TRIM(wine_combined!B2:M6498))` entered in cell B2.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Wine_type	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality							
2	White	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6							
3	White	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6							
4	White	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6							
5	White	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6							
6	White	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6							
7	White	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6							
8	White	6.2	0.32	0.16	7	0.045	30	136	0.9949	3.18	0.47	9.6	6							
9	White	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6							
10	White	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6							
11	White	8.1	0.22	0.43	1.5	0.044	28	129	0.9938	3.22	0.45	11	6							
12	White	8.1	0.27	0.41	1.45	0.033	11	63	0.990	2.99	0.56	12	5							
13	White	8.6	0.23	0.4	4.2	0.035	17	109	0.9947	3.14	0.53	9.7	5							
14	White	7.9	0.18	0.37	1.2	0.04	16	75	0.992	3.18	0.63	10.8	5							
15	White	6.6	0.16	0.4	1.5	0.044	48	143	0.9912	3.54	0.52	12.4	7							
16	White	8.3	0.42	0.62	19.25	0.04	41	172	1.0002	2.98	0.67	9.7	5							
17	White	6.6	0.17	0.38	1.5	0.032	28	112	0.9914	3.25	0.55	11.4	7							
18	White	6.3	0.48	0.04	1.1	0.046	30	99	0.9928	3.24	0.36	9.6	6							
19	White	6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8							
20	White	7.4	0.34	0.42	1.1	0.033	17	171	0.9917	3.12	0.53	11.3	6							
21	White	6.5	0.31	0.14	7.5	0.044	34	133	0.9955	3.22	0.5	9.5	5							
22	White	6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8							
23	White	6.4	0.31	0.38	2.9	0.038	17	141	0.9961	3.38	0.53	10.3	6							
24	White	6.8	0.26	0.42	1.7	0.047	17	132	0.9914	3.18	0.49	11.6	6							
25	White	7.2	0.32	0.36	2	0.033	37	114	0.9906	3.1	0.71	12.3	7							
26	White	8.5	0.24	0.39	10.4	0.044	20	142	0.9974	3.2	0.53	10	6							

Screenshot of Microsoft Excel showing the formula bar with the formula `=TRIM(wine_combined!A2:A6498)` entered in cell A2.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Wine_type	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality						
2	White	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6						
3	White	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6						
4	White	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6						
5	White	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6						
6	White	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6						
7	White	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6						
8	White	6.2	0.32	0.16	7	0.045	30	136	0.9949	3.18	0.47	9.6	6						
9	White	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6						
10	White	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6						
11	White	8.1	0.22	0.43	1.5	0.044	28	129	0.9938	3.22	0.45	11	6						
12	White	8.1	0.27	0.41	1.45	0.033	11	63	0.990	2.99	0.56	12	5						
13	White	8.6	0.23	0.4	4.2	0.035	17	109	0.9947	3.14	0.53	9.7	5						
14	White	7.9	0.18	0.37	1.2	0.04	16	75	0.992	3.18	0.63	10.8	5						
15	White	6.6	0.16	0.4	1.5	0.044	48	143	0.9912	3.54	0.52	12.4	7						
16	White	8.3	0.42	0.62	19.25	0.04	41	172	1.0002	2.98	0.67	9.7	5						
17	White	6.6	0.17	0.38	1.5	0.032	28	112	0.9914	3.25	0.55	11.4	7						
18	White	6.3	0.48	0.04	1.1	0.046	30	99	0.9928	3.24	0.36	9.6	6						
19	White	6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8						
20	White	7.4	0.34	0.42	1.1	0.033	17	171	0.9917	3.12	0.53	11.3	6						
21	White	6.5	0.31	0.14	7.5	0.044	34	133	0.9955	3.22	0.5	9.5	5						
22	White	6.2	0.66	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8						
23	White	6.4	0.31	0.38	2.9	0.038	19	102	0.9912	3.17	0.35	11	7						
24	White	6.8	0.26	0.42	1.7	0.049	41	122	0.993	3.47	0.48	10.5	8						
25	White	7.6	0.67	0.14	1.5	0.074	25	168	0.9937	3.05	0.51	9.3	5						
26	White	6.6	0.27	0.41	1.3	0.052	16	142	0.9951	3.42	0.47	10	6						

7. Create an attribute with equally sized bins for the ‘free sulfur dioxide’ attribute. You may choose the number of bins.

Result:

8. Create a categorical attribute based on the column ‘quality’. This new categorical attribute should consist of 3 levels: ‘bad’, ‘good’ and ‘best’ defined as follows:
quality of less than 4 – ‘bad’

quality between 4 and 7 – ‘good’

quality greater than 7 – ‘best’

Result:

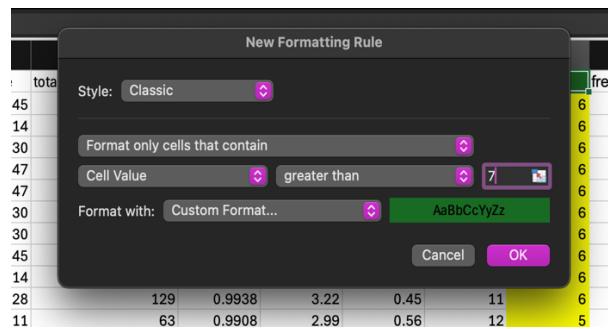
	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	oxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	free_sulphur_bins						
2	45	170	1.001	3	0.45	8.8	6		4	good				
3	14	132	0.994	3.3	0.49	9.5	6		1	good				
4	30	97	0.9951	3.26	0.44	10.1	6		3	good				
5	47	186	0.9956	3.19	0.4	9.9	6		5	good				
6	47	186	0.9956	3.19	0.4	9.9	6		5	good				
7	30	97	0.9951	3.26	0.44	10.1	6		3	good				
8	30	136	0.9949	3.18	0.47	9.6	6		3	good				
9	45	170	1.001	3	0.45	8.8	6		4	good				
10	14	132	0.994	3.3	0.49	9.5	6		1	good				
11	28	129	0.9938	3.22	0.45	11	6		3	good				
12	11	63	0.9908	2.99	0.56	12	5		1	good				
13	17	109	0.9947	3.14	0.53	9.7	5		2	good				
14	16	75	0.992	3.18	0.63	10.8	5		2	good				
15	48	143	0.9912	3.54	0.52	12.4	7		5	good				
16	41	172	1.0002	2.98	0.67	9.7	5		4	good				
17	28	112	0.9914	3.25	0.55	11.4	7		3	good				
18	30	99	0.9928	3.24	0.36	9.6	6		3	good				
19	29	75	0.9892	3.33	0.39	12.8	8		3	best				
20	17	171	0.9917	3.12	0.53	11.3	6		2	good				
21	34	133	0.9955	3.22	0.5	9.5	5		4	good				
22	29	75	0.9892	3.33	0.39	12.8	8		3	best				
23	19	102	0.9912	3.17	0.35	11	7		2	good				
24	41	122	0.993	3.47	0.48	10.5	8		4	best				
25	25	168	0.9937	3.05	0.51	9.3	5		3	good				
26	16	142	0.9951	3.42	0.47	10	6		2	good				
27	56	245	0.9955	3.25	0.5	10.4	6		5	good				
28	35	146	0.993	3.45	0.44	10	6		4	good				
29	32	141	0.9961	3.38	0.53	10.5	6		3	good				
30	17	132	0.9914	3.19	0.49	11.6	6		2	good				

9. Use the internet or Gen AI to do some research on Excel's Conditional Formatting.

Now, use conditional formatting to color the row in green if the quality of wine is greater than 7. If the quality of wine is lesser than 7, color the row yellow.

NOTE: You must use conditional formatting and show screenshots for how you have achieved the result. MacOS users may face some issues while attempting this. Create the conditional formatting rule regardless – you will be graded on the creation of the rule.

Result:





File: winequality-white.xlsx

Home Insert Draw Page Layout Formulas Data Review View Automate

M1

	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	free_sulphur_bins	quality_category		
2	.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6	4	good	
3	.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6	1	good	
4	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6	3	good	
5	.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6	5	good	
6	.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6	5	good	
7	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6	3	good	
8	.16	7	0.045	30	136	0.9949	3.18	0.47	9.6	6	3	good	
9	.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6	4	good	
10	.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6	1	good	
11	.43	1.5	0.044	28	129	0.9938	3.22	0.45	11	6	3	good	
12	.41	1.45	0.033	11	63	0.9908	2.99	0.56	12	5	1	good	
13	0.4	4.2	0.035	17	109	0.9947	3.14	0.53	9.7	5	2	good	
14	.37	1.2	0.04	16	75	0.992	3.18	0.63	10.8	5	2	good	
15	0.4	1.5	0.044	48	143	0.9912	3.54	0.52	12.4	7	5	good	
16	.62	19.25	0.04	41	172	1.0002	2.98	0.67	9.7	5	4	good	
17	.38	1.5	0.032	28	112	0.9914	3.25	0.55	11.4	7	3	good	
18	.04	1.1	0.046	30	99	0.9928	3.24	0.36	9.6	6	3	good	
19	.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	6	3	best	
20	.42	1.1	0.033	17	171	0.9917	3.12	0.53	11.3	6	2	good	
21	.14	7.5	0.044	34	133	0.9955	3.22	0.5	9.5	5	4	good	
22	.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	6	3	best	
23	.38	2.9	0.038	19	102	0.9912	3.17	0.35	11	7	2	good	
24	.42	1.7	0.049	41	122	0.993	3.47	0.48	10.5	6	4	best	
25	.14	1.5	0.074	25	168	0.9937	3.05	0.51	9.3	5	3	good	
26	.41	1.3	0.052	16	142	0.9951	3.42	0.47	10	6	2	good	
27	.32	9	0.046	56	245	0.9955	3.25	0.5	10.4	6	5	good	
28	.35	1	0.052	35	146	0.993	3.45	0.44	10	6	4	good	
29	.39	8.7	0.051	32	141	0.9961	3.38	0.53	10.5	6	3	good	

10. Gen AI Prompt Engineering: use your favorite AI Chat Bot to perform same work we did in class on boston_housing dataset, i.e clean data, manipulate data, analyze data, and viz data. While doing this task, you may need to subscribe to one of the bots, e.g. Chat GPT.

Solution:

I am using Perplexity.ai

Prompt:

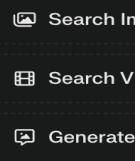
The screenshot shows the Perplexity.ai interface. At the top, there is a header bar with the user's name 'sinhasim3321' and a timestamp '21m'. Below the header, a main prompt reads: 'Use the given spreadsheet and perform data cleanup in the excel'. Underneath the prompt, there is a 'Pro Search' button and a 'Sources' section. The 'Sources' section contains a single item: 'boston_hosuing-lab01.xlsx'. On the right side of the interface, there are three buttons: 'Search Image', 'Search Video', and 'Generate'.

Solution by AI:

The screenshot shows the Perplexity.ai generated solution for data cleanup. It starts with a heading 'Perplexity' and a brief introduction: 'To clean up the data in the given Excel spreadsheet, we'll need to perform several steps to ensure the data is properly formatted and ready for analysis. Here's a step-by-step guide on how to clean up the data:'. Below this, there is a section titled 'Data Cleanup Process' which lists five steps:

- 1. Remove unnecessary rows**
 - Delete the first 22 rows containing the description and variable explanations.
 - Keep only the header row and the actual data.
- 2. Create proper column headers**
 - Replace the current header row with clear, concise column names based on the variable descriptions.
 - Use the following headers: CRIM, ZN, INDUS, CHAS, NOX, RM, AGE, DIS, RAD, TAX, PTRATIO, B, LSTAT, MEDV
- 3. Format data types**
 - Ensure all numeric columns are formatted as numbers.
 - Format the CHAS column as a binary integer (0 or 1).
- 4. Remove any blank rows or columns**
 - Scan the spreadsheet for any empty rows or columns and delete them.
- 5. Check for and handle missing values**

- Scan the spreadsheet for any empty rows or columns and delete them.
- 5. Check for and handle missing values**
- Look for any cells with missing data (e.g., blank cells or error values).
 - Decide whether to remove rows with missing data or impute values based on the analysis requirements.
- 6. Trim whitespace**
- Use Excel's TRIM function to remove any leading or trailing spaces in text fields.
- 7. Check for and handle outliers**
- Use Excel's sorting and filtering features to identify any extreme values that may be outliers.
 - Investigate these values and decide whether to keep, modify, or remove them based on the context of the data.
- 8. Standardize text entries**
- Ensure consistent capitalization and spelling in any text columns (although in this dataset, most columns are numeric).
- 9. Remove duplicates**
- Use Excel's "Remove Duplicates" feature to eliminate any duplicate rows, if present.

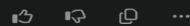


- 10. Create a data dictionary**
- Add a new sheet to the workbook.
 - Create a table with columns for Variable Name, Description, and Data Type.
 - Fill in the information for each variable based on the original descriptions.
- 11. Freeze panes and apply filters**
- Freeze the top row to keep headers visible while scrolling.
 - Apply filters to all columns for easy data exploration.
- 12. Save a clean version**
- Save the cleaned dataset as a new file to preserve the original data.



By following these steps, you'll have a clean, well-structured dataset ready for analysis. The cleaned data will be easier to work with and less likely to produce errors or misleading results in subsequent analyses.

[Share](#) [Rewrite](#)



Excel:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	CINN	ZN	INDUS	NOX	RPM	AGE	DIST	RAD	TAX	PRATIO	B	LSTAT	Medv																	
2	0.49032	16	2.31	0	0.638	8.675	65.2	4.09	1	296	16.3	394.0	4.98	24.																
3	0.02731	0	7.07	0	0.649	8.425	76.1	4.09	1	247	16.3	399.0	9.34	24.	0.0932	0	7.23	0	0.469	6.499	16.2	2.98	15.3	296.9	8.88	21.6	0.00000			
4	0.02729	0	7.07	0	0.469	8.4671	61.1	4.0971	2	242	17.8	390.83	4.93	34.7	0.07731	0	7.67	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.6	9.54	21.6	0.00000	
5	0.02727	0	2.18	0	0.469	8.4671	54.2	6.0622	3	223	18.7	396.0	5.33	36.2	0.09327	0	7.97	0	0.469	6.421	41.3	4.9671	2	242	17.8	396.6	9.54	21.6	0.00000	
6	0.02985	0	2.18	0	0.458	7.147	64.3	58.7	6.0622	3	223	18.7	394.12	5.21	28.7	0.09005	0	7.18	0	0.458	7.147	49.8	4.6022	3	223	18.7	394.63	2.84	33.4	0.00000
7	0.02989	12.5	7.07	0	0.524	8.034	0.93	0.622	5	311	15.2	396.0	15.6	23.8	0.09289	0	7.18	0	0.524	6.012	66.6	5.5605	5	311	15.2	395.6	5.33	36.2	0.00000	
8	0.14455	12.5	7.07	0	0.524	8.152	96.1	5.9505	5	311	15.2	396.0	19.15	27.1	0.08829	0	7.87	0	0.524	6.012	66.6	5.5605	5	311	15.2	395.6	12.49	22.9	0.00000	
9	0.17044	12.5	7.07	0	0.524	8.034	85.9	6.5921	5	311	15.2	396.0	17.1	20.8	0.02124	0	7.87	0	0.524	6.531	100	6.0821	5	311	15.2	386.63	29.93	16.5	0.00000	
10	0.17044	12.5	7.07	0	0.524	8.034	94.3	6.3467	5	311	15.2	396.0	20.8	16.0	0.07004	0	7.87	0	0.524	6.531	100	6.0821	5	311	15.2	386.63	17.1	18.8	0.00000	
11	0.17044	12.5	7.07	0	0.524	8.034	89.9	6.2267	5	311	15.2	396.0	15.71	18.9	0.02489	0	7.87	0	0.524	6.537	95.3	6.1131	5	311	15.2	386.63	17.1	18.8	0.00000	
12	0.09579	12.5	7.07	0	0.524	8.034	39	5.4509	5	311	15.2	396.0	15.71	21.7	0.11747	0	7.87	0	0.524	6.009	82.9	6.2367	5	311	15.2	396.6	12.37	18.9	0.00000	
13	0.09579	12.5	7.07	0	0.524	8.034	39	5.4509	5	311	15.2	396.0	15.71	21.7	0.03337	0	7.87	0	0.524	6.009	82.9	6.2367	5	311	15.2	396.6	12.37	18.9	0.00000	
14	0.09579	12.5	7.07	0	0.524	8.034	39	5.4509	5	311	15.2	396.0	15.71	21.7	0.03337	0	7.87	0	0.524	6.009	82.9	6.2367	5	311	15.2	396.6	12.37	18.9	0.00000	
15	0.09579	12.5	7.07	0	0.524	8.034	39	5.4509	5	311	15.2	396.0	15.71	21.7	0.03337	0	7.87	0	0.524	6.009	82.9	6.2367	5	311	15.2	396.6	12.37	18.9	0.00000	
16	0.09579	12.5	7.07	0	0.524	8.034	39	5.4509	5	311	15.2	396.0	15.71	21.7	0.03337	0	7.87	0	0.524	6.009	82.9	6.2367	5	311	15.2	396.6	12.37	18.9	0.00000	
17	0.09579	12.5	7.07	0	0.524	8.034	39	5.4509	5	311	15.2	396.0	15.71	21.7	0.03337	0	7.87	0	0.524	6.009	82.9	6.2367	5	311	15.2	396.6	12.37	18.9	0.00000	
18	0.09579	12.5	7.07	0	0.524	8.034	39	5.4509	5	311	15.2	396.0	15.71	21.7	0.03337	0	7.87	0	0.524	6.009	82.9	6.2367	5	311	15.2	396.6	12.37	18.9	0.00000	
19	0.09579	12.5	7.07	0	0.524	8.034	39	5.4509	5	311	15.2	396.0	15.71	21.7	0.03337	0	7.87	0	0.524	6.009	82.9	6.2367	5	311	15.2	396.6	12.37	18.9	0.00000	
20	0.09579	12.5	7.07	0	0.524	8.034	39	5.4509	5	311	15.2	396.0	15.71	21.7	0.03337	0	7.87	0	0.524	6.009	82.9	6.2367	5	311	15.2	396.6	12.37	18.9	0.00000	
21	0.7558	0	8.14	0	0.524	8.034	67.95	3.7965	4	307	21	390.95	11.28	18.2	0.08271	0	8.14	0	0.524	5.456	36.6	3.7965	4	307	21	388.89	11.69	20.2	0.00000	
22	0.09579	0	8.14	0	0.524	8.034	5.5	3.7965	4	307	21	390.95	11.28	18.2	0.08271	0	8.14	0	0.524	5.456	36.6	3.7965	4	307	21	388.89	11.69	20.2	0.00000	
23	0.09579	0	8.14	0	0.524	8.034	5.5	3.7965	4	307	21	390.95	11.28	18.2	0.08271	0	8.14	0	0.524	5.456	36.6	3.7965	4	307	21	388.89	11.69	20.2	0.00000	
24	0.09579	0	8.14	0	0.524	8.034	5.5	3.7965	4	307	21	390.95	11.28	18.2	0.08271	0	8.14	0	0.524	5.456	36.6	3.7965	4	307	21	388.89	11.69	20.2	0.00000	
25	0.09579	0	8.14	0	0.524	8.034	5.5	3.7965	4	307	21	390.95	11.28	18.2	0.08271	0	8.14	0	0.524	5.456	36.6	3.7965	4	307	21	388.89	11.69	20.2	0.00000	
26	0.09579	0	8.14	0	0.524	8.034	5.5	3.7965	4	307	21	390.95	11.28	18.2	0.08271	0	8.14	0	0.524	5.456	36.6	3.7965	4	307	21	388.89	11.69	20.2	0.00000	
27	0.09579	0	8.14	0	0.524	8.034	5.5	3.7965	4	307	21	390.95	11.28	18.2	0.08271	0	8.14	0	0.524	5.456	36.6	3.7965	4	307	21	388.89	11.69	20.2	0.00000	
28	0.09579	0	8.14	0	0.524	8.034	5.5	3.7965	4	307	21	390.95	11.28	18.2	0.08271	0	8.14	0	0.524	5.456	36.6	3.7965	4	307	21	388.89	11.69	20.2	0.00000	
29	0.09579	0	8.14	0	0.524	8.034	5.5	3.7965	4	307	21	390.95	11.28	18.2	0.08271	0	8.14	0	0.524	5.456	36.6	3.7965	4	307	21	388.89	11.69	20.2	0.00000	
30	0.09579	0	8.14	0	0.524	8.034	5.5	3.7965	4	307	21	390.95	11.28	18.2	0.08271	0	8.14	0	0.524	5.456	36.6	3.7965	4	307	21	388.89	11.69	20.2	0.00000	
31	0.09579	0	8.14	0	0.524	8.034	6.674	8.73	4	307	21	390.23	11.98	21	0.07299	0	8.14	0	0.524	6.495	94.4	4	307	21	387.94	12.8	18.4	0.00000		
32	0.09579	0	8.14	0	0.524	8.034	6.674	8.73	4	307	21	390.23	11.98	21	0.07299	0	8.14	0	0.524	6.495	94.4	4	307	21	387.94	12.8	18.4	0.00000		
33	1.35472	0	8.14	0	0.524	8.034	6.674	8.73	4	307	21	376.73	13.04	14.5	1.13081	0	8.14	0	0.524	5.713	94.1	4	307	21	360.17	23.6	13.7	0.00000		
34	0.09579	0	8.14	0	0.524	8.034	6.674	8.73	4	307	21	376.73	13.04	14.5	1.13081	0	8.14	0	0.524	5.713	94.1	4	307	21	360.17	23.6	13.7	0.00000		
35	0.15172	0	8.14	0	0.524	8.034	5.703	9.6	3.7872	4	307	21	358.77	18.35	13.1	1.38799	0	8.14	0	0.524	5.85	82	4	307	21	232.6	27.77	13.2	0.00000	
36	0.15172	0	8.14	0	0.524	8.034	5.703	9.6	3.7872	4	307	21	358.77	18.35	13.1	1.38799	0	8.14	0	0.524	5.85	82	4	307	21	232.6	27.77	13.2	0.00000	
37	0.09579	0	8.14	0	0.524	8.034	5.703	9.6	3.7872	4	307	21	358.77	18.35	13.1	1.38799	0	8.14	0	0.524	5.85	82	4	307	21	232.6	27.77	13.2	0.00000	
38	0.09579	0	8.14	0	0.524	8.034	5.703	9.6	3.7872	4	307	21	358.77	18.35	13.1	1.38799	0	8.14	0	0.524	5.85	82	4	307	21	232.6	27.77	13.2	0.00000	
39	0.09579	0	8.14	0	0.524	8.034	5.703	9.6	3.7872	4	307	21	358.77	18.35	13.1	1.38799	0	8.14	0	0.524	5.85	82	4	307	21	232.6	27.77	13.2	0.00000	
40	0.09579	0	8.14	0	0.524	8.034	5.703	9.6	3.7872	4	307	21	358.77	18.35	13.1	1.38799	0	8.14	0	0.524	5.85	82	4	307	21	232.6	27.77	13.2	0.00000	
41	0.09579	0	8.14	0	0.524	8.034	5.703																							

Solution by AI:

To perform data manipulation on the Book3.xlsx file you've uploaded, follow these steps in Excel:

1. Open and Prepare the Data

1. Open Book3.xlsx in Excel.
2. Ensure you're working on the "data_manipulation" sheet.

2. Remove Unnecessary Columns

1. Delete the "DIST" column as it appears to be a misspelling of "DIS".
2. Select column H (DIST).
3. Right-click and choose "Delete".

3. Rename Columns

1. Rename "CRM" to "CRIM" for consistency.
2. Double-click on cell A1 and change "CRM" to "CRIM".

4. Format Data Types

1. Select all data (excluding headers).
2. Go to Home > Number Format > Number.
3. Set decimal places to 3 for consistency.

5. Handle Missing Values

1. Press Ctrl+F to open the Find dialog.
2. Search for blank cells and fill them with the average of the column:
 - a. Select the entire column with blanks.
 - b. Go to Home > Editing > Find & Select > Go To Special > Blanks.
 - c. Enter this formula: =AVERAGE(column_letter:column_letter)
 - d. Press Ctrl+Enter to fill all blank cells.
 - e. Copy and paste values to replace formulas.

6. Create New Features

1. In column O, add header "ROOMS_PER_DWELLING".
2. In O2, enter formula: =G2
3. In column P, add header "POPULATION_DENSITY".
4. In P2, enter formula: =H2/I2
5. In column Q, add header "HIGH_TAX".
6. In Q2, enter formula: =IF(K2>MEDIAN(\$K\$2:\$K\$507),1,0)
7. Copy formulas down for all rows.

7. Normalize Numerical Features

1. Select columns A:N (excluding headers).
2. Go to Data > Data Analysis > Descriptive Statistics.
3. Check "Summary statistics" and "Output Range", select an empty cell.

b. Repeat for each numerical column.

8. Encode Categorical Variables

1. For CHAS column:
 - a. Insert a column next to CHAS.
 - b. Use formula: =IF(D2=1,"Yes","No")
 - c. Copy down for all rows.

9. Create Interaction Terms

1. In a new column, add header "RM_LSTAT_INTERACTION".
2. Use formula: =G2*N2
3. In another new column, add header "NOX_INDUS_INTERACTION".
4. Use formula: =E2*C2

10. Bin Continuous Variables

1. For AGE column:
 - a. Insert a new column next to AGE.
 - b. Use formula: =IF(H2<25,"Young",IF(H2<50,"Middle",IF(H2<75,"Senior","Old")))

11. Calculate Correlations

1. Select all numerical columns.
2. Go to Data > Data Analysis > Correlation.
3. Select an output range in your worksheet.

Excel:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
2	0.00632	18	2.31	0	0.538	6.575	65.2	4.09	1	296	15.3	396.9	4.98	24
3	0.02731	0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.9	9.14	21.6

Prompt:

The RM Column which is average number of houses per dwelling
i want to perform data manipulation on it by creating an attribute and bins for the data

Pro Search

2 sources

Sources

Book3.xlsx

boston_housing-lab01.xlsx

Book3

boston_housing-lab01

Solution by AI:

To perform data manipulation on the RM column (average number of rooms per dwelling) and create an attribute with bins, we'll follow these steps:

1. Analyze the RM column
2. Create bins
3. Assign categories to the bins
4. Add a new column for the binned data

1. Analyze the RM column

First, let's look at the range of values in the RM column:

Minimum value: 3.561
Maximum value: 8.780

2. Create bins

We'll create 5 equal-width bins for the RM values:

1. 3.561 to 4.605
2. 4.606 to 5.649
3. 5.650 to 6.693
4. 6.694 to 7.737
5. 7.738 to 8.780

3. Assign categories to the bins

We'll assign descriptive categories to each bin:

1. Very Small (3.561 to 4.605)
2. Small (4.606 to 5.649)
3. Medium (5.650 to 6.693)
4. Large (6.694 to 7.737)
5. Very Large (7.738 to 8.780)

4. Add a new column for the binned data

To create the new attribute in Excel, follow these steps:

1. Add a new column next to the RM column, and name it "ROOM_SIZE_CATEGORY".
2. In the first cell of the new column (assuming RM is in column F), enter this formula:

```
text
=IF(F2<4.606, "Very
Small", IF(F2<5.65, "Small", IF(F2<6.694, "Medium", IF(F2<7.738, "Large", "Very
Large"))))
```

3. Copy this formula down for all rows in the dataset.

This formula will categorize each RM value into one of the five bins we created.

Excel:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	RM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PIRATIO	B	LSTAT	MEDV	Bin_RM	Age_category					
2	0.00632	18	2.31	0	0.538	6.575	65.2	4.09	1	296	15.3	396.9	4.98	24	Medium	Mature					
3	0.02731	0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.9	9.14	21.6	Medium	Historic					
4	0.02729	0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7	Large	Mature					
5	0.03237	0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4	Large	ModeratelyAged					
6	0.06905	0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.9	5.33	36.2	Large	Mature					
7	0.06905	0	2.18	0	0.458	6.43	56.7	6.0622	3	222	18.7	394.12	5.21	28.7	Medium	Mature					
8	0.08829	12.5	7.87	0	0.524	6.012	66.8	5.9505	5	311	15.2	395.6	12.5	22.0	Medium	Mature					
9	0.14455	12.5	7.87	0	0.524	6.172	96.1	5.9505	5	311	15.2	396.9	19.15	27.1	Medium	Historic					
10	0.2124	12.5	7.87	0	0.524	5.631	109	6.0821	5	311	15.2	386.63	29.98	16.5	Small	Historic					
11	0.17004	12.5	7.87	0	0.524	6.004	85.9	6.5921	5	311	15.2	386.71	17.1	18.9	Medium	Historic					

Prompt:

Now give me categorical attribute based on the column 'age'

Pro Search

2 sources

Sources

Book3.xlsx

boston_housing-lab01.xlsx

Book3

boston_housing-lab01

Solution by AI:

Let's create the proper categories for the AGE column (proportion of owner-occupied units built prior to 1940):

1. In a new column next to AGE, name it "AGE_CATEGORY"
2. Use this formula to properly categorize the values:

text

```
=IF(H2<=25,"New Construction",IF(H2<=50,"Moderately Aged",IF(H2<=75,"Mature",IF(H2<=100,"Historic",""))))
```



This will create four meaningful categories:

- New Construction: 0-25%
- Moderately Aged: 25.1-50%
- Mature: 50.1-75%
- Historic: 75.1-100%

The formula will now correctly categorize the values and you won't see all entries marked as "New". This better reflects the actual distribution of building ages in your dataset, where many properties have high AGE values indicating older construction.

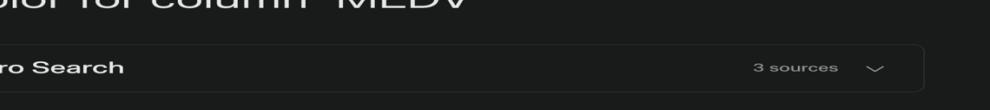
To verify the categorization worked correctly, you can create a PivotTable to see the distribution of properties across these age categories.

Excel:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	RIM	ZN	INDUS	CHAS	NOX	RM	6.575	65.2	4.09	1	296	15.2	396.9	4.98	24	Bin_RM	Age_category			
2	0.06821	18	2.31	0	0.538					2	242	17.8	396.9	9.14	21.6	Medium	Mature			
3	0.02731	0	7.07	0	0.469	6.421	78.9	4.9671		2	242	17.8	392.83	4.03	34.7	Medium	Historic			
4	0.02729	0	7.07	0	0.469	7.185	61.1	4.9671		2	242	17.8	392.83			Large	Mature			
5	0.03237	0	2.18	0	0.458	6.998	45.8	6.0622		3	222	18.7	394.63	2.94	33.4	Large	ModeratelyAged			
6	0.06905	0	2.18	0	0.458	7.147	54.2	6.0622		3	222	18.7	396.9	5.33	36.2	Large	Mature			
7	0.02985	0	2.18	0	0.458	6.43	58.7	6.0622		3	222	18.7	394.12	5.21	28.7	Medium	Mature			
8	0.08829	12.5	7.87	0	0.524	6.012	66.6	5.5605		5	311	15.2	395.6	12.43	22.9	Medium	Mature			
9	0.14455	12.5	7.87	0	0.524	6.172	96.1	5.9505		5	311	15.2	396.9	19.15	27.1	Medium	Historic			
10	0.21124	12.5	7.87	0	0.524	5.631	100	6.0821		5	311	15.2	386.63	29.93	16.5	Small	Historic			
11	0.17004	12.5	7.87	0	0.524	6.004	85.9	6.5921		5	311	15.2	386.71	17.1	18.9	Medium	Historic			
12	0.22489	12.5	7.87	0	0.524	6.377	94.3	6.3467		5	311	15.2	392.52	20.45	15	Medium	Historic			
13	0.11747	12.5	7.87	0	0.524	6.009	82.9	6.2267		5	311	15.2	396.9	13.27	18.9	Medium	Historic			
14	0.09378	12.5	7.87	0	0.524	5.889	39	5.4509		5	311	15.2	390.5	15.71	21.7	Medium	ModeratelyAged			
15	0.62976	0	8.14	0	0.538	5.949	61.8	4.7075		4	307	21	396.9	8.26	20.4	Medium	Mature			
16	0.63796	0	8.14	0	0.538	6.096	84.5	4.4619		4	307	21	380.02	10.26	18.2	Medium	Historic			
17	0.62739	0	8.14	0	0.538	5.834	56.5	4.4986		4	307	21	395.62	8.47	19.9	Medium	Mature			

Prompt:

use book 3 and perform conditional formatting to color for column "MEDV"



Pro Search

3 sources

Sources

Book3.xlsx

	MEDV
Small	New
Medium	New
Medium	New
Large	New
Large	New
Large	New
Medium	New
Medium	New
Medium	New
Small	New
Medium	New
Small	New

Book3.xlsx

Book3

Solution by AI:

 **Perplexity**

I'll help you create conditional formatting for the MEDV (Median value of owner-occupied homes) column in Book3. Here's how to do it:

1. Select the MEDV column (last column in your dataset)
2. Go to Home > Conditional Formatting > Color Scales
3. Apply the following rules:
 - Low values (below 20): Red
 - Mid values (20-35): Yellow
 - High values (35 and above): Green

To set this up in Excel:

1. Select the MEDV column
2. Home > Conditional Formatting > New Rule
3. Choose "Format all cells based on their values"
4. Select "3-Color Scale"
5. Set the following values:
 - Minimum: Type = Number, Value = 0, Color = Red
 - Midpoint: Type = Number, Value = 25, Color = Yellow
 - Maximum: Type = Number, Value = 50, Color = Green
 - Maximum: Type = Number, Value = 50, Color = Green

This will create a gradient where:

- Houses with lower values (≤ 20) appear in red shades
- Mid-range houses (20-35) appear in yellow shades
- Premium houses (≥ 35) appear in green shades

The color gradient will help visualize the distribution of house values across your dataset, making it easier to identify patterns and outliers in the housing prices.

Excel

The screenshot shows the 'New Formatting Rule' dialog box. At the top, it says 'Style: 3-Color Scale'. Below that, there are three sections: 'Minimum', 'Midpoint', and 'Maximum'. Each section has a 'Type:' dropdown set to 'Number', a 'Value:' input field, and a 'Color' picker. The 'Value' for Minimum is 0 (red), Midpoint is 25 (yellow), and Maximum is 50 (green). At the bottom right are 'Cancel' and 'OK' buttons.

	J	K	L	M	N	O	P	Q
TAX		PTRATIO	B	LSTAT	MEDV			
1	296	15.3	396.9	4.98	24			
2	242	17.8	396.9	9.14	21.6			
2	242	17.8	392.83	4.03	34.7			
3	222	18.7	394.63	2.94	33.4			
3	222	18.7	396.9	5.33	36.2			
3	222	18.7	394.12	5.21	28.7			
5	311	15.2	395.6	12.43	22.9			
5	311	15.2	396.9	19.15	27.1			
5	311	15.2	386.63	29.93	16.5			
5	311	15.2	386.71	17.1	18.9			
5	311	15.2	392.52	20.45	15			
5	311	15.2	396.9	13.27	18.9			
5	311	15.2	390.5	15.71	21.7			
4	307	21	396.9	8.26	20.4			
4	307	21	380.02	10.26	18.2			
4	307	21	395.62	8.47	19.9			
4	307	21	386.85	6.58	23.1			
4	307	21	386.75	14.67	17.5			
4	307	21	288.99	11.69	20.2			
4	307	21	390.95	11.28	18.2			
4	307	21	376.57	21.02	13.6			
4	307	21	392.53	13.83	19.6			
4	307	21	396.9	18.72	15.2			
4	307	21	394.54	19.88	14.5			
4	307	21	394.33	16.3	15.6			
4	307	21	303.42	16.51	13.9			
4	307	21	376.88	14.81	16.6			
4	307	21	306.38	17.28	14.8			
4	307	21	387.94	12.8	18.4			
4	307	21	380.23	11.98	21			
4	307	21	360.17	22.6	12.7			
4	307	21	376.73	13.04	14.5			
4	307	21	232.6	27.71	13.2			
4	307	21	358.77	18.35	13.1			
4	307	21	248.31	20.34	13.5			
5	279	19.2	396.9	9.68	18.9			
5	279	19.2	377.56	11.41	20			
5	279	19.2	396.9	8.77	21			
5	279	19.2	393.43	10.13	24.7			
3	252	18.3	395.63	4.32	30.8			
3	252	18.3	395.62	1.98	34.9			

Prompt:

This is my excel after manipulation perform data analysis. Give me the steps to perform on my excel

The screenshot shows a search interface with a dark background. At the top, there's a search bar with the placeholder "Pro Search" and a dropdown menu showing "3 sources". Below the search bar, there's a section titled "Sources" with two entries: "Book3.xlsx" and "Book4.xlsx". Each entry has a small thumbnail preview and a "Book3" or "Book4" label.

Solution by AI:

Perplexity

Based on your Excel data, I'll provide steps to perform data analysis on your manipulated dataset:

1. Descriptive Statistics

- Calculate basic statistics for numerical columns:
 - Select Data > Data Analysis > Descriptive Statistics
 - Input Range: Select all numerical columns
 - Output Range: Choose a new location in your worksheet
 - Check "Summary statistics"

Excel:

	C20	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1		Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12											
2																								
3	an	0.00632	Mean	18	Mean	2.31	Mean	0	Mean	0.538	Mean	6.575	Mean	65.2	Mean	4.09	Mean	1	Mean	296	Mean	15.3	Mean	
4	andard Error	0	Standard Err	0.538	Standard Err	6.575	Standard Err	65.2	Standard Err	0	Standard Err													
5	ean	0.00632	Median	18	Median	2.31	Median	0	Median	0.538	Median	6.575	Median	65.2	Median	4.09	Median	1	Median	296	Median	15.3	Median	
6	nus	#NA!	#NA!	#NA!	#NA!	#NA!	#NA!	#NA!	#NA!	#NA!	#NA!	#NA!	#NA!	#NA!	#NA!									
7	andard Deviation	#DIV/0!	Standard Dev	#DIV/0!	Standard Dev	#DIV/0!	Standard Dev	#DIV/0!	Standard Dev	#DIV/0!	Standard Dev	#DIV/0!	Standard Dev	#DIV/0!	Standard Dev									
8	ple Variance	#DIV/0!	Sample Vari	#DIV/0!	Sample Vari	#DIV/0!	Sample Vari	#DIV/0!	Sample Vari	#DIV/0!	Sample Vari	#DIV/0!	Sample Vari	#DIV/0!	Sample Vari									
9	ross	#DIV/0!	Kurtosis	#DIV/0!	Kurtosis	#DIV/0!	Kurtosis	#DIV/0!	Kurtosis	#DIV/0!	Kurtosis	#DIV/0!	Kurtosis	#DIV/0!	Kurtosis									
10	ewness	#DIV/0!	Skewness	#DIV/0!	Skewness	#DIV/0!	Skewness	#DIV/0!	Skewness	#DIV/0!	Skewness	#DIV/0!	Skewness	#DIV/0!	Skewness									
11	age	0	Range	0	Range	0	Range	0	Range	0	Range	0	Range	0	Range									
12	sum	0.00632	Minimum	18	Minimum	2.31	Minimum	0	Minimum	0.538	Minimum	6.575	Minimum	65.2	Minimum	4.09	Minimum	1	Minimum	296	Minimum	15.3	Minimum	
13	imum	0.00632	Maximum	18	Maximum	2.31	Maximum	0	Maximum	0.538	Maximum	6.575	Maximum	65.2	Maximum	4.09	Maximum	1	Maximum	296	Maximum	15.3	Maximum	
14	m	0.00632	Sum	18	Sum	2.31	Sum	0	Sum	0.538	Sum	6.575	Sum	65.2	Sum	4.09	Sum	1	Sum	296	Sum	15.3	Sum	
15	nt	1	Count	1	Count	1	Count	1	Count	1	Count	1	Count	1	Count									
16																								
17																								

Prompt:

give steps for data visualization in excel using book_4 spreadsheet

Pro Search

10 sources

Sources

Book3.xlsx

Book4.xlsx

Show all

Solution by AI:

1. Create a Histogram for MEDV (Median Home Values)

1. Select the MEDV column
2. Go to Insert > Insert Statistic Chart > Histogram
3. Customize the bins:
 - Right-click on x-axis
 - Select Format Axis
 - Choose "Bin Width" = 5
 - This will show the distribution of house prices in \$5,000 intervals

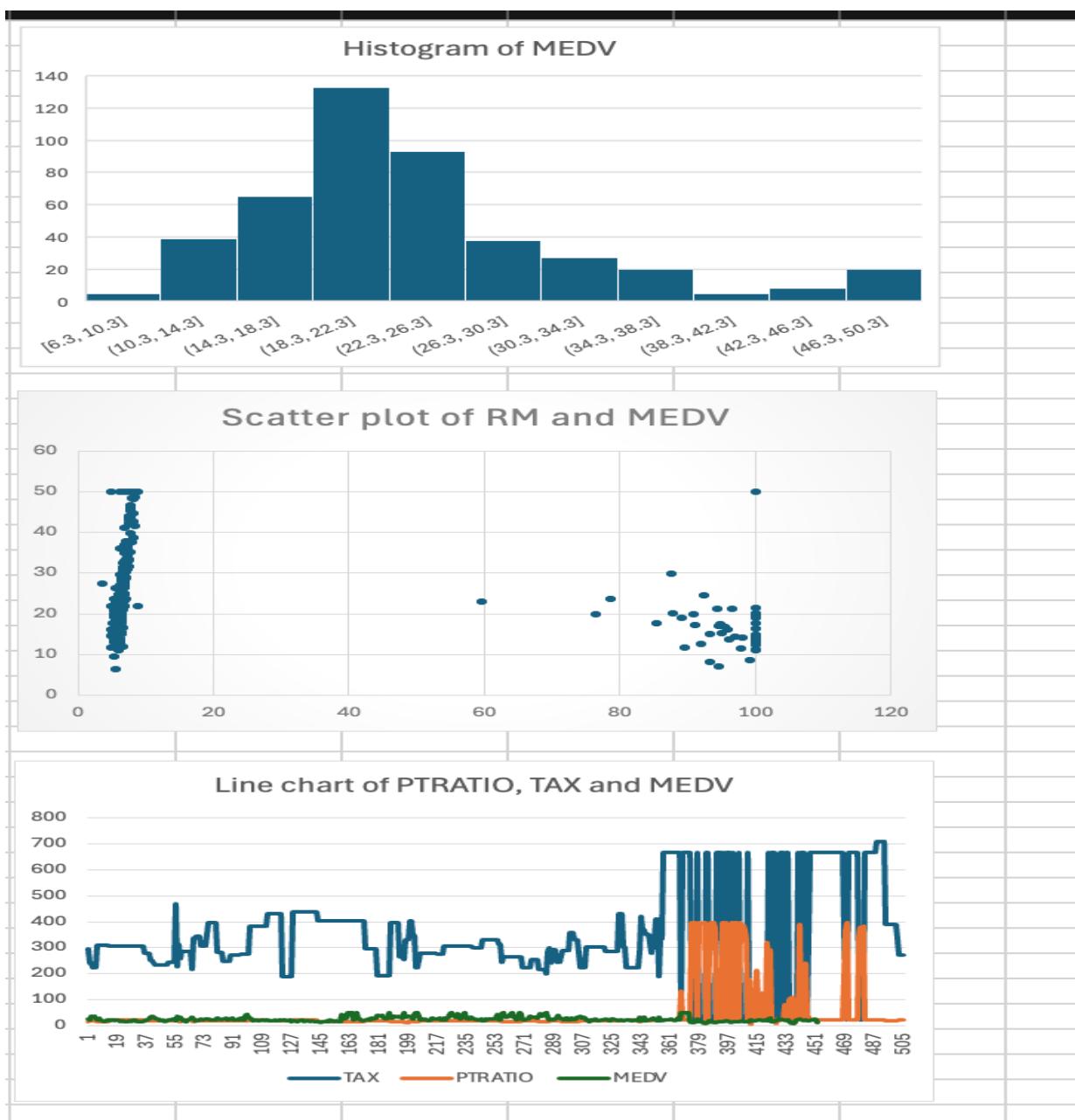
2. Create a Scatter Plot for Price vs Room Size

1. Select both RM (average rooms) and MEDV columns
2. Go to Insert > Scatter Chart
3. Customize the chart:
 - Add chart title: "House Prices vs Room Size"
 - X-axis label: "Average Number of Rooms"
 - Y-axis label: "Median Value (\$1000s)"

6. Create a Multiple Line Chart

1. Select MEDV, TAX, and PTRATIO columns
2. Insert > Line Chart
3. Format with:
 - Title: "Price, Tax, and Pupil-Teacher Ratio Trends"
 - Different line styles for each variable
 - Legend for clear identification

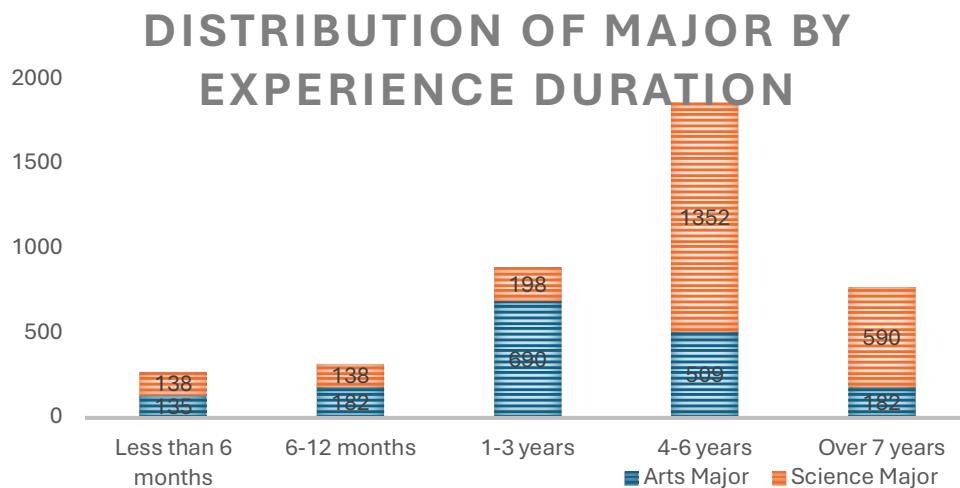
Excel:



Lab 02**PART I**

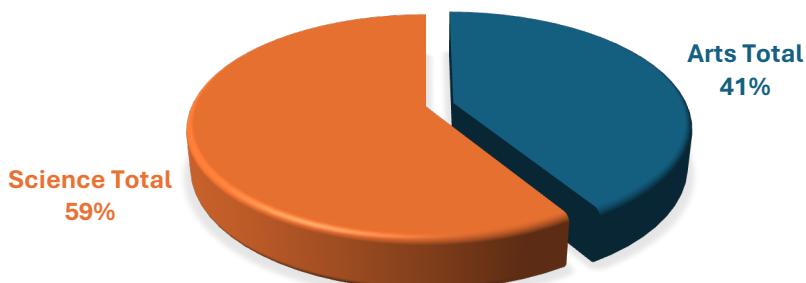
Use the “Internet Technology.xlsx” dataset. Create a worksheet and do the following:

1. Create a stacked-column chart for the five time periods for both Arts Majors and Science Majors. Let Time on the Internet be your X-axis category label and put Major in the legend. Move the legend to the bottom of the chart. Include an appropriate chart title and label your axes appropriately.



2. Create a 3-D pie chart of the total data (combining majors). Include an appropriate title. Remove the legend and display data labels for the category name and percentage. Create the chart on the worksheet and move it below the first chart. Rotate the pie to display the largest slice at the front. Explode this one slice of the pie

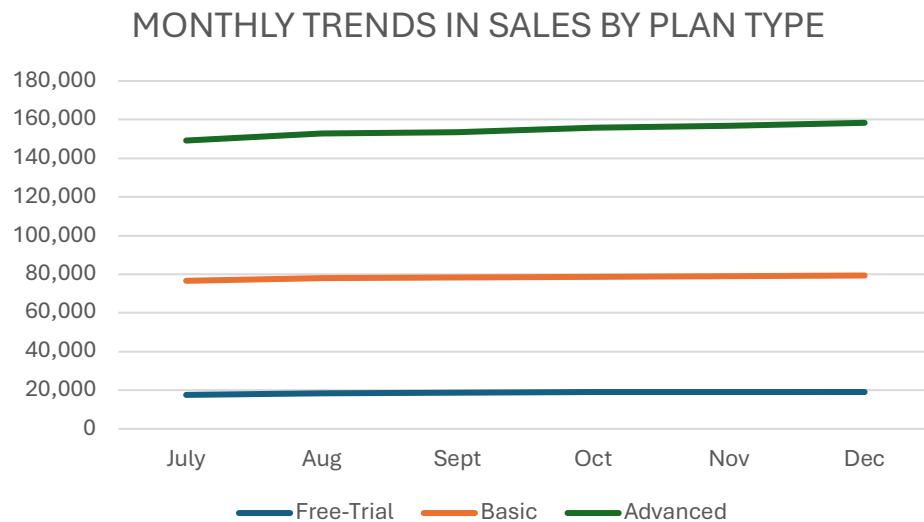
OVERALL DISTRIBUTION OF ARTS AND SCIENCE MAJORS



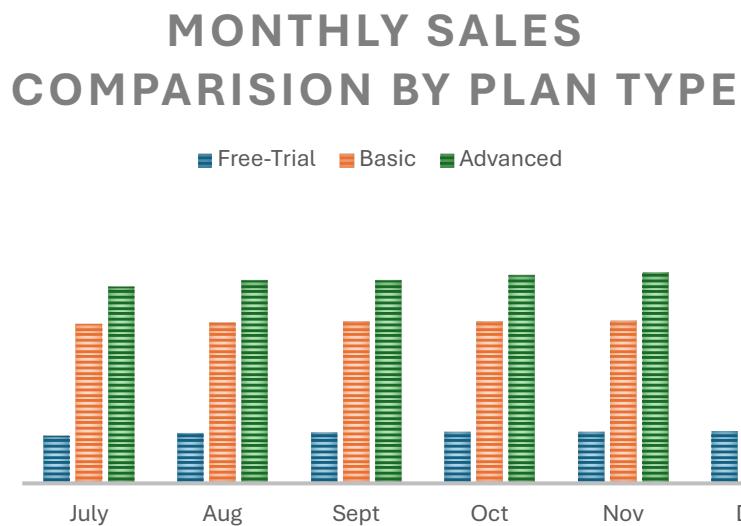
PART II

Use the “Antivirus Sales.xlsx” dataset. Create a worksheet and do the following:

1. Create a Line Chart with each line representing the Free-trial, basic and advanced.



2. Create a Grouped Bar chart with a bars for Free-trial, basic and advanced.



3. Create sparklines for each of the categories as seen in class

	A	B	C	D	E	F	G	H
1		July	Aug	Sept	Oct	Nov	Dec	Spark_Lines
2	Free-Trial	17,589	18,540	18,690	18,982	18,990	19,082	
3	Basic	59,025	59,395	59,778	59,823	60,012	60,284	
4	Advanced	72,566	74,905	75,001	76,934	77,904	78,992	
5								

PART III Gen AI

I am using Perplexity

1. Use the StateSales sheet in the “more charts.xlsx” file to create the pareto chart we did in class. Ask the bot to explain the 80/20 rule in the resulting chart.

What does it mean.

Prompt:

Create a pareto chart for the data in the given sheet

Pro Search 4 sources ▾

Sources

Book5.xlsx	Book3.xlsx	Book4.xlsx
Book5	Book3	Book4

The screenshot shows a search interface with a dark background. At the top, there is a search bar with the placeholder "Create a pareto chart for the data in the given sheet". Below the search bar, there is a "Pro Search" button and a "4 sources" dropdown menu. Under the "Sources" heading, there are three items: "Book5.xlsx", "Book3.xlsx", and "Book4.xlsx". The "Book3.xlsx" item is highlighted with a green box around its thumbnail and name. The thumbnail for Book3.xlsx shows a table with columns "Size" and "Type", containing rows like "Large", "Medium", and "Small". A small number "3" is visible at the bottom left of the thumbnail, indicating it is the third item in the list.

Solution by AI:

⌘ Perplexity

I'll help you create a Pareto chart using the state sales data from Book5. Here are the detailed steps:

1. Prepare the Data

1. Sort the sales data in descending order:
 - Select both State and Sales columns
 - Go to Data > Sort
 - Sort by Sales column in descending order

The sorted data should look like this:

- Illinois: 105
- Ohio: 60
- Pennsylvania: 45
- Wisconsin: 27
- Missouri: 21
- Colorado: 12
- Indiana: 9
- Kansas: 6

Steps:

1. Calculate individual percentages:

```
text
=(B2/SUM($B$2:$B$9))*100
```



2. Calculate cumulative percentages:

```
text
=SUM($C$2:C2)
```



Where:

- Column B contains your frequency/count values
- Column C contains individual percentages
- Column D will contain cumulative percentages

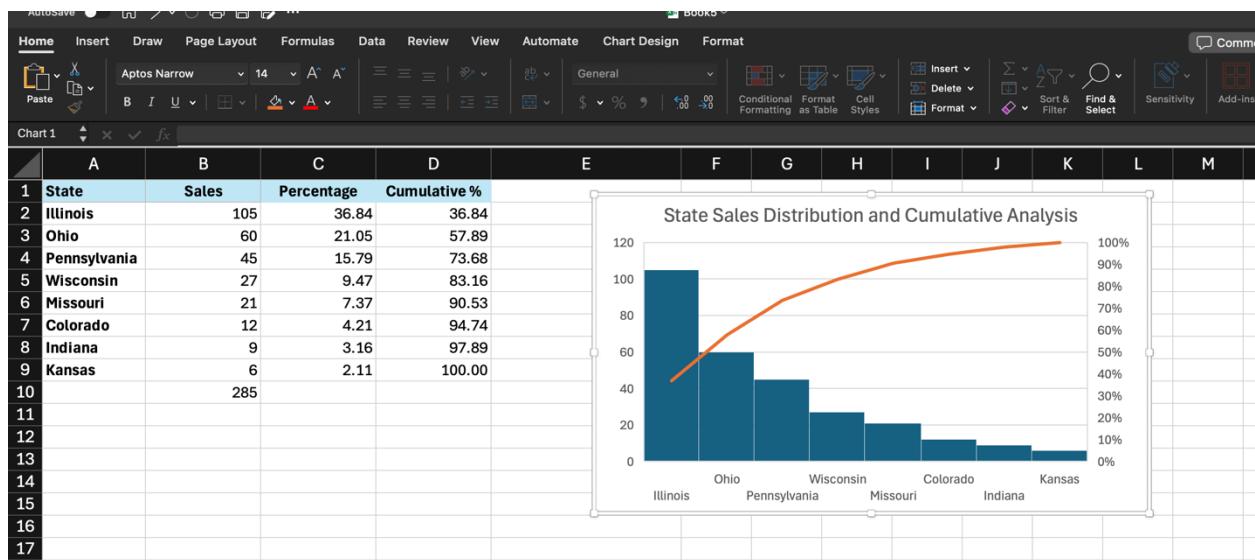
Example layout:

State	Sales	Percentage	Cumulative %
Illinois	105	=105/285*100	=C2
Ohio	60	=60/285*100	=C2+C3
Pennsylvania	45	=45/285*100	=SUM(\$C\$2:C4)

3. Create the Pareto Chart

1. Select your prepared data
2. Go to Insert > Combo Chart
3. Choose:
 - Columns for Sales values
 - Line for Cumulative Percentage
4. Format the chart:
 - Left Y-axis: Sales values (0-120)
 - Right Y-axis: Percentages (0-100%)
 - X-axis: State names
 - Add title: "State Sales Pareto Analysis"

Excel:



explain the 80/20 rule in the resulting chart.
Explain what does it mean

The Pareto chart shown demonstrates the 80/20 principle (Pareto Principle), but in this case, it doesn't exactly follow the classic 80/20 distribution. Let's analyze the actual distribution:

Sales Distribution Analysis

The cumulative percentage line shows that:

- Illinois and Ohio together (top 2 states) account for approximately 58% of total sales (165 out of 285 units)
- Adding Pennsylvania (third state) brings the cumulative total to about 74% of sales
- The remaining 5 states (Wisconsin, Missouri, Colorado, Indiana, and Kansas) only contribute about 26% of total sales

Business Implications

This distribution reveals that:

- The business is heavily dependent on three key states
- Illinois alone contributes 36.8% of total sales
- The bottom three states (Colorado, Indiana, and Kansas) collectively contribute less than 10% of total sales

While this doesn't exactly match the classic 80/20 rule (where 80% of effects come from 20% of causes), it still demonstrates the principle of uneven distribution, where a small number of states account for the majority of sales. This insight can be valuable for:

- Resource allocation
- Sales territory planning

The chart helps identify where the business should focus its efforts for maximum impact

While this doesn't exactly match the classic 80/20 rule (where 80% of effects come from 20% of causes), it still demonstrates the principle of uneven distribution, where a small number of states account for the majority of sales. This insight can be valuable for:

- Resource allocation
- Sales territory planning
- Market development strategies
- Risk management

The chart helps identify where the business should focus its efforts for maximum impact on overall sales performance.

2. Use the StateSalesLong sheet in the “more charts.xlsx” file to create the Box & Whisker (Box) chart we did in class. Ask the bot to explain the insight of the resulting chart, including outliers. Ask the bot why use Box chart for better visualization

Prompt:

use the StateSales sheet in the Book5 spreadsheet and create a box and whisker (box) chart
Using book 5 only perform the above task

Solution by AI:



To create a Box and Whisker plot for the sales data in Book5:

1. Select the Sales column data (105, 60, 45, 27, 21, 12, 9, 6)
2. Go to Insert > Statistical Charts > Box and Whisker
3. Format the chart:
 - Title: "Sales Distribution Analysis"
 - Y-axis label: "Sales Volume"

The resulting Box and Whisker plot will show:

Excel:

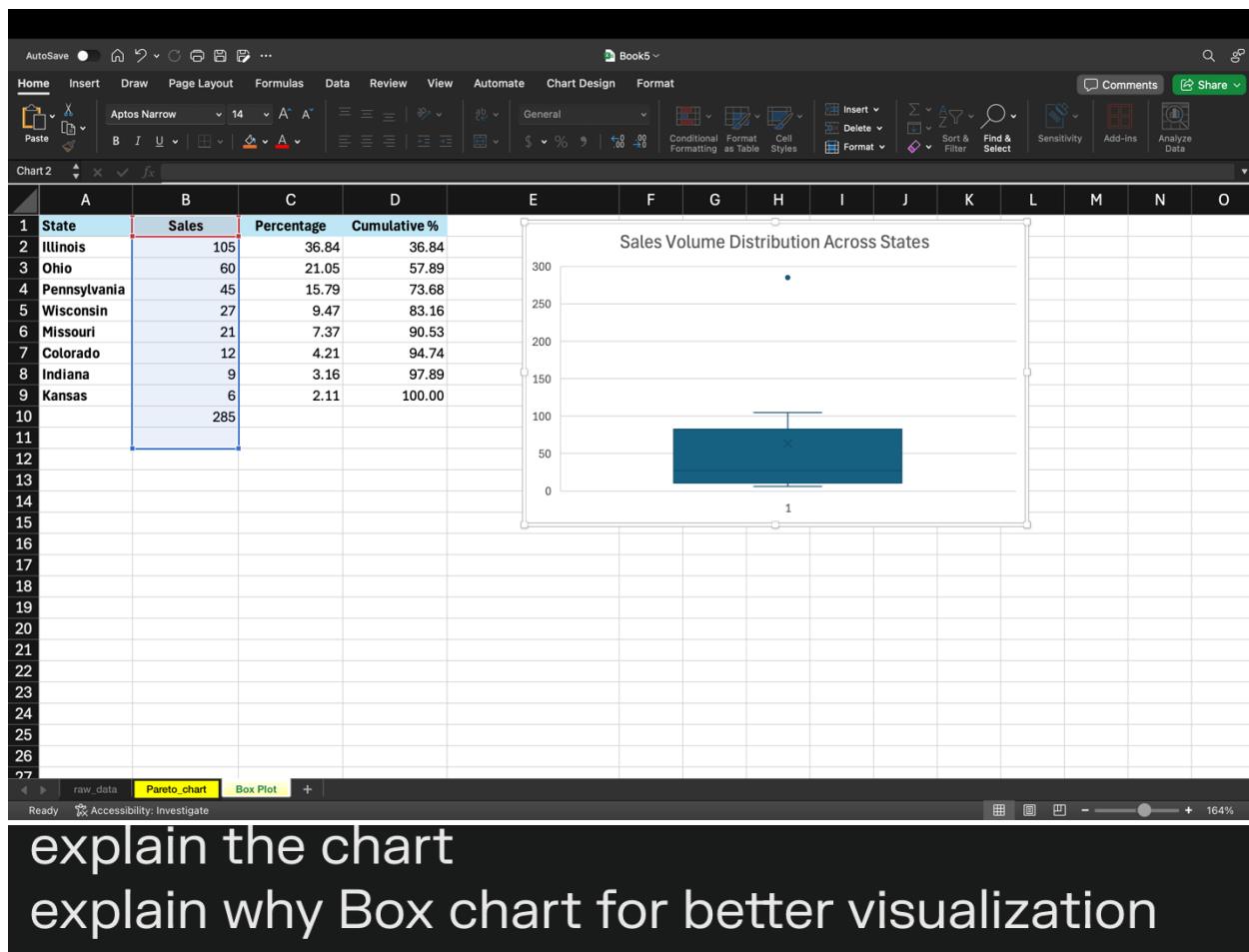


Chart Explanations

Pareto Chart Analysis:

- Shows sales distribution across 8 states with cumulative percentage line
- Illinois leads with 105 units (36.8% of total sales)
- Top 3 states (Illinois, Ohio, Pennsylvania) account for ~74% of total sales
- Clear visualization of sales concentration in few key states
- Orange line shows cumulative percentage reaching 100% across all states

Box Plot Analysis:

- Median sales: 24 units (middle line in box)
- Interquartile Range (IQR): ~10.5 to ~56.25 units (box boundaries)
- Outlier: Illinois at 105 units (dot above box)
- Whiskers extend from minimum (6) to maximum non-outlier value
- Shows significant spread in sales performance

Why Box Plot is Better for Visualization

1. **Statistical Summary:** Provides five-number summary (min, Q1, median, Q3, max) in one compact view
2. **Outlier Detection:** Clearly identifies Illinois as an outlier in sales performance
3. **Distribution Shape:** Shows right-skewed distribution of sales data
4. **Data Spread:** Effectively displays the variation and central tendency of sales
5. **Quick Insights:** Enables immediate understanding of sales distribution without needing exact numbers

3. Use the Transactions sheet in Gen_AI_Prompt_Engineering.xlsx and ask Gen AI bot to categorize each transaction as debt or credit. Also ask the bot to create a bar chart to visualize the data

Prompt:

using book6 spreadsheet categorize the transactions as debit or credit for Transactions mentioned.
also create a bar chart to visualize the data in excel

Edit Query

Solution by AI:

Here's an easy way to categorize transactions as debit or credit in Excel:

1. Add a new column called "Transaction Type"
2. Use this simple IF formula based on transaction amount:

text

```
=IF(B2>0, IF(OR(ISNUMBER(SEARCH("Deposit",A2)), ISNUMBER(SEARCH("Transfer from",A2)), ISNUMBER(SEARCH("Refund",A2)), ISNUMBER(SEARCH("Return",A2))), "Credit", "Debit"), "Credit")
```

To sum up credit and debit values in Excel, follow these steps:

Using SUM Function for Credits and Debits

1. For Credits:

text

```
=SUMIF(Category_Column, "Credit", Amount_Column)
```

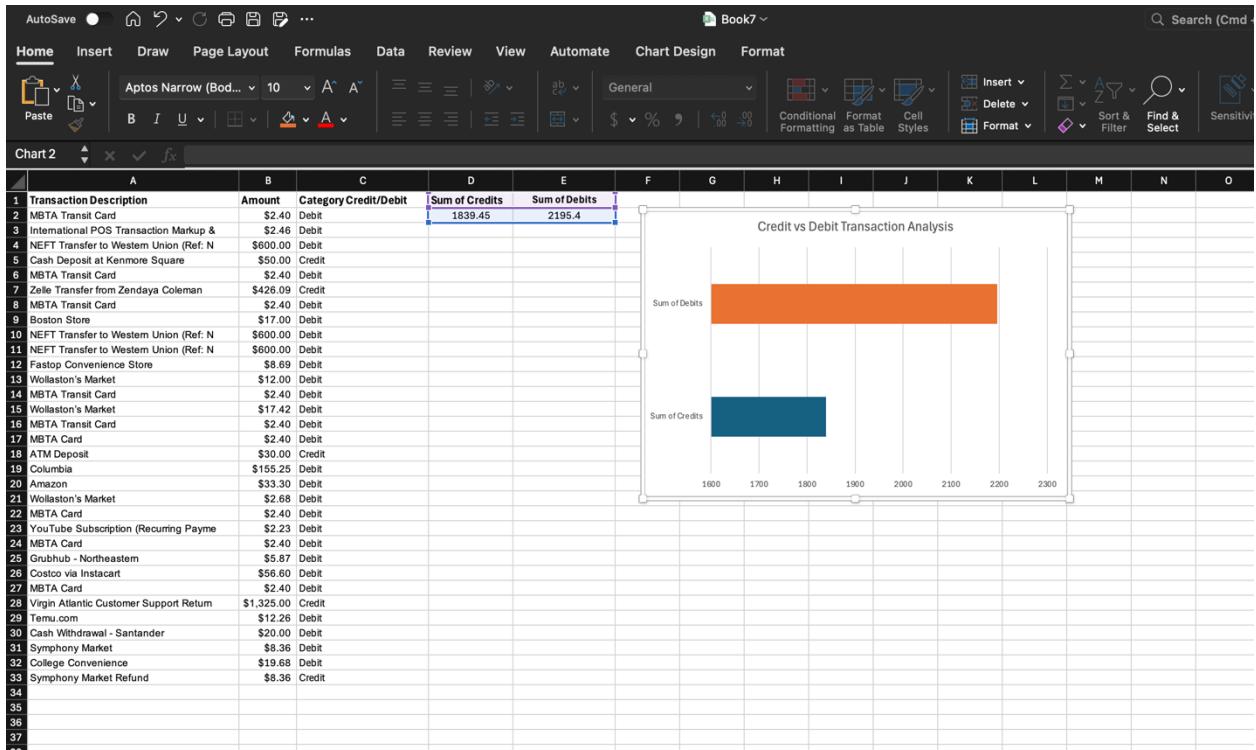
2. For Debits:

text

```
=SUMIF(Category_Column, "Debit", Amount_Column)
```



Excel:



4. Use the Restockings sheet in Gen_AI_Prompt_Engineering.xlsx and ask Gen AI bot to use its data to calculate the Recommended Restocking Quantity and display in in the Recommended Restocking Quantity column. Also ask the bot to create a bar chart to visualize the data. Also, copy the results to excel and submit file.

Prompt:

calculate recommended restocking quantity.
also create a bar chart to visualie

Solution by AI:

I'll help you calculate the recommended restocking quantity for each product and create a bar chart visualization.

Calculate Recommended Restocking Quantity

Use this formula in Excel for the Recommended Restocking Quantity column:

text

```
= (Average_Monthly_Sales * Seasonality_Factor) - Current_Inventory
```

Example calculations:

- Lays Chips: $(120 * 1.1) - 100 = 32$
- Coca-Cola Soda: $(150 * 1.2) - 120 = 60$
- Hershey's Chocolate: $(100 * 0.9) - 80 = 10$

Create Bar Chart

- After calculating all recommended quantities:

- Select the Product Name and Recommended Restocking Quantity columns
- Go to Insert > Bar Chart
- Choose "Clustered Bar" type

Excel:

