ASSIGNMENT 4

1. **Product Inventory Cleanup**

Data Set: Inventory Records

Task:

- Identify and remove any incomplete or erroneous entries in the inventory records.
- Check for inconsistencies in text.
- Standardize the date column.
- Create 3 new columns titled year, month, and weekday. This will be gotten from the date column.
- Apply conditional formatting to highlight any products with negative stock levels or zero unit prices.

Product ID	Product Name	Category	Price ($)	Quantity	Supplier	Purchase Date
101	Apple MacBook Pro	Electronic	2000	5	Tech Giant	2022-10-15
102	Nike Air Max Shoes	Footwear	150	10	Foot Locker	2022-09-20
103	Samsung Galaxy S21	Electronics	900	8	Best Buy	2022-11-05
104	HP Pavilion Laptop	Electronics	NA	-3	Office Depot	2022-12-01
105	Adidas Originals Hoodie	Apparel	50	15	Adidas Store	2022-10-10

Product ID	Product Name	Category	Price ($)	Quantity	Supplier	Purchase Date
106	Sony PlayStation 5	Electronic	500	6	GameStop	2022-11-20
107	Dell Inspiron Desktop	Electronics	700	4	Walmart	09/15/22
108	Levi's 501 Jeans	Apprel	80	12	Levi's Store	2022-10-30
109	Apple iPhone 13	Electronics	1000	-7	Apple Store	2022-12-05
110	Reebok Nano X Shoes	Footwear	120	9	Reebok Outlet	2022-09-25
111	ASUS ROG Gaming Laptop	Electronics	1500	2	Best Buy	2022-11-10
112	Calvin Klein T-Shirt	Apparel	30	20	Macy's	2022-10-05
113	Sony WH-1000XM4 Headphones	Electronics	300	4	Best Buy	2022-12-20
114	Timberland 6-Inch Boots	Footwear	180	6	Timberland Store	2022-09-30
115	Fitbit Charge 5	Electronics	150	-8	Fitbit Store	2022-11-15
116		Electronics	900	3		2022/08/22
117	Adidas Originals Hoodie	Clothing	50	15	Adidas Store	2022-10-10
118	Google Pixel 6	Electronics	800	10	Best Buy	2022-07-05
119			-200	4	Unknown	2022-06-30
120	Apple Watch Series 7	Electronics	600	5	Apple Store	Not Available

2. **Data: Sales**

   Task:

   - Calculate “Month”, “Weekday” columns from the date column. Make the weekday column to be in ‘Mon’, ’Tue’, ’Wed’…format.
   - Calculate new profit and profit per quantity columns.
   - Calculate a new column called ‘Profit percentage’, this is calculated by (profit/cost of good)*100.
   - Calculate new cost per quantity and price per quantity columns.
   - Apply conditional formatting to highlight transactions with a profit margin above the average threshold.
   - Apply a function to categorize each transaction as "High", "Medium", or "Low" based on its profit. Call the column ‘Transaction status’.
   - Use a function to create a new column combining the product name and transaction ID. Call the column ‘Product_ID’.
   - Use the IF function to classify transactions with a profit percentage above 50% as "High profit by %" and below 30% as "Low Profit by %". Call the column ‘Transaction by profit %’.

Transaction ID	Date	Product	Cost of Good ($)	Price Sold ($)	Quantity
101	2023-02-15	Laptop	700	1000	3
102	2023-01-20	Smartphone	400	600	5

Transaction ID	Date	Product	Cost of Good ($)	Price Sold ($)	Quantity
103	2023-03-08	Headphones	50	100	10
104	2023-04-10	Tablet	300	450	2
105	2023-01-12	Smartwatch	150	200	4
106	2023-03-25	Speaker	80	120	6
107	2023-02-05	Monitor	200	300	3
108	2023-04-01	Keyboard	30	50	8
109	2023-03-18	Mouse	20	40	12
110	2023-01-30	Router	80	100	5
111	2023-02-17	Printer	150	200	3
112	2023-03-11	Webcam	50	80	4
113	2023-04-03	External HDD	120	180	2
114	2023-01-28	USB Drive	10	20	20
115	2023-03-01	Power Bank	30	50	5

Transaction ID	Date	Product	Cost of Good ($)	Price Sold ($)	Quantity
116	2023-01-05	Headset	70	100	6
117	2023-02-10	Tablet Case	20	40	3
118	2023-03-14	Laptop Bag	50	80	4
119	2023-04-20	Wireless Mouse	25	50	2
120	2023-02-22	Printer Ink	40	60	2
121	2023-01-18	External SSD	150	250	1
122	2023-03-05	HDMI Cable	10	15	5
123	2023-04-15	USB Hub	15	25	4
124	2023-01-08	Mouse Pad	5	10	10
125	2023-02-14	Phone Case	15	30	3
126	2023-03-20	Ethernet Cable	8	15	2
127	2023-04-25	Laptop Stand	20	40	1
128	2023-01-03	Keyboard Cleaner	5	10	6

Transaction ID	Date	Product	Cost of Good ($)	Price Sold ($)	Quantity
129	2023-02-19	Screen Protector	10	20	3
130	2023-04-08	Surge Protector	25	40	2

3. **Data: Health.**

   Task:

   - What is the correlation between weight and cholesterol level?
   - Is there any significant correlation between height and blood sugar level?
   - How well can we predict cholesterol level based on weight?
   - Can we estimate cholesterol level based on height?
   - Does age have a correlation with cholesterol level?
   - Create a new column 'systolic', this is derived from the first 3 values of the blood pressure.
   - Is there a correlation between systolic and cholesterol level?
   - Is there a linear relationship between weight and blood sugar level? If yes, Provide the linear regression model and the R-square
   - Can we predict blood sugar level based on age? If yes, What will be the blood sugar level of a 70 year old?

Serial No.	Age (years)	Height (cm)	Weight (kg)	Blood Pressure (mmHg)	Cholesterol Level (mg/dL)	Blood Sugar Level (mg/dL)
1	35	175	70	120/80	200	90

Serial No.	Age (years)	Height (cm)	Weight (kg)	Blood Pressure (mmHg)	Cholesterol Level (mg/dL)	Blood Sugar Level (mg/dL)
2	42	168	75	130/85	210	95
3	28	180	80	125/82	190	88
4	55	163	68	140/90	230	105
5	48	172	73	128/84	220	98
6	38	160	65	118/78	180	85
7	62	178	77	135/88	240	110
8	29	167	72	122/81	200	92
9	45	175	79	130/85	210	95
10	50	170	75	128/82	215	100
11	33	172	70	125/80	190	90
12	40	163	68	120/78	205	98
13	57	178	80	132/86	225	105
14	36	169	73	128/84	195	88
15	47	173	75	135/88	210	100
16	31	166	70	118/76	185	90
17	52	180	82	140/90	230	105
18	44	168	76	130/85	215	98
19	30	171	72	125/82	200	92
20	58	175	78	132/86	220	100
21	39	170	74	120/80	205	95
22	49	173	76	128/84	210	98

4. **Data: Environment.**

   Task:

   - Is there a correlation between temperature and energy consumption? Show in a chart
   - What is the correlation between humidity and solar radiation?
   - Can we predict energy consumption based on temperature using linear regression?
   - What is the regression equation for predicting energy consumption based on wind speed?
   - Do pressure and wind speed have any relationship?
   - What is the correlation between solar radiation and energy consumption?

D	Temperature (°C)	Humidity (%)	Pressure (hPa)	Wind Speed (km/h)	Solar Radiation (W/m^2)	Energy Consumption (kWh)
1	20	65	1015	10	200	30
2	22	70	1010	12	180	32
3	25	68	1008	15	150	35
4	18	60	1012	8	220	28
5	23	72	1005	13	190	33
6	19	63	1014	9	210	29
7	21	67	1011	11	175	31
8	24	75	1007	14	160	34
9	17	58	1013	7	230	27
10	26	73	1003	16	140	36
11	16	55	1016	6	240	26
12	27	78	1002	17	130	37

D	Temperature (°C)	Humidity (%)	Pressure (hPa)	Wind Speed (km/h)	Solar Radiation (W/m^2)	Energy Consumption (kWh)
13	15	50	1017	5	250	25
14	28	80	1001	18	120	38
15	14	45	1018	4	260	24
16	29	82	1000	19	110	39
17	13	40	1019	3	270	23
18	30	85	999	20	100	40
19	12	35	1020	2	280	22
20	31	88	998	21	90	41
21	11	30	1021	1	290	21
22	32	90	997	22	80	42

5. **Dataset: Student test scores and study habits**

Tasks:

- Is there a correlation between study hours and test scores? Plot.
- What is the correlation between practice tests taken and test scores? Plot.
- Is there a correlation between attendance percentage and test scores? Plot.
- What is the correlation between sleep hours per night and test scores? Plot.
- Is there a correlation between stress level and test scores? Plot.
- Can we use study hours per week to predict test scores? If so, how strong is the relationship?
- How well can practice tests taken predict test scores? Provide the linear regression model and the R-square.

- Can attendance percentage be used to predict test scores? If so, what is the regression equation?
- Is sleep hours per night a significant predictor of test scores?
- How well does stress level predict test scores, and is the relationship statistically significant?
- Calculate correlation coefficients between test scores and study hours, sleep time, extracurricular activities, etc.

Student ID	Study Hours (per week)	Practice Tests Taken	Test Score (%)	Attendance (%)	Sleep Hours (per night)	Stress Level (1-10)
101	10	4	85	95	7	3
102	8	3	78	90	6	5
103	12	5	92	98	8	2
104	6	2	70	85	5	7
105	9	4	80	92	7	4
106	11	5	88	96	8	3
107	7	3	75	88	6	6
108	10	4	82	93	7	4
109	8	3	76	88	6	5
110	12	6	90	97	8	3
111	9	4	80	91	7	4
112	10	5	85	94	7	3
113	8	3	78	89	6	5
114	11	6	88	95	8	3
115	7	3	74	87	6	6
116	9	4	81	92	7	4

Student ID	Study Hours (per week)	Practice Tests Taken	Test Score (%)	Attendance (%)	Sleep Hours (per night)	Stress Level (1-10)
117	10	5	84	94	7	3
118	8	3	77	89	6	5
119	12	6	91	97	8	3
120	9	4	79	91	7	4
121	11	5	86	95	8	3
122	7	3	73	88	6	6