Milestone Project 1: Scenario 1

Overview

You have been hired by a space research company named AWAY (Aliens Where Are You). One of their biggest projects has been to compile a list of yellow-dwarf stars. Yellow-dwarf stars are important because that is the type of star our sun is. AWAY scientists have theorized that other yellow-dwarf stars may be able to support life the way our sun supports life on earth. AWAY looks for planets around these stars in the search for alien lifeforms.

You have been tasked with determining which stars are most likely to have nearby planets that could possibly support life.

Audience: Scientists

About the Dataset

- **Temperature:** The average temperature of the star
- **L:** The L column stands for “luminosity,” which measures the brightness of a star
- **R:** The R column stands for “radius,” which is the distance from the center of a star to its outer edge
- **Is_star:** Displays 1 if the object is a star or 0 if the object is not a star
- **A_M:** The column A_M means “absolute magnitude,” which is the magnitude of a star when measured from a distance of 10 parsecs (1 parsec = 3.26 lightyears)
- **Spectral Class:** The group that a star belongs to depending on its spectrum and luminosity

- **Color:** The color of the star

CLASS	COLOR	TEMPERATURE
O	Blue	>= 30,000 K
B	Blue-White	10,000-30,000 K
A	White	7,500-10,000 K
F	Yellow-White	6,000-7,500 K
G	Yellow	5,200-6,000 K
K	Orange	3,700-5,200 K
M	Red	2,400-3,700 K

- **Type:** Type of star

NUMBER	TYPE
0	Red Dwarf
1	Brown Dwarf
2	White Dwarf
3	Main Sequence
4	Super Giants
5	Hyper Giants

Part 1 - Data Preparation

Tool: Microsoft Excel

Step 1: Define the goal.

What is the goal for this data analysis? What questions are you trying to answer?

determining which stars are most likely to have nearby planets
that could possibly support life.

Step 2: Remove irrelevant columns.

Not every column will be useful for analysis. We can remove any columns where every row is the same. *Hint: An easy way to tell if every row is the same is to use the filter columns tool.*

What column(s) did you remove and why?

I would have removed Is_Star but it was not in the data set as stated

Step 3: Identify typos.

Search for any typos that exist in the dataset in the *Color* column and correct them.

Additionally, ensure that the data in the *Color* column is formatted to be the same. For example, two-worded colors have a hyphen (-) between the words, while others do not. Choose one consistent format. Used =SUBSTITUTE(A1,"-","") to remove hyphens and =Proper for caps

Step 4: Identify nulls and missing values.

Almost every dataset contains missing values. The goal is to handle such values uniformly throughout the dataset.

How did you choose to handle missing values and why?

Step 5: Remove duplicates.

Remove any duplicates in the dataset. Navigate to the **Data** tab in the Data Tools section and click the **Remove Duplicates** button.

Part 2 - Data Exploration

Tool: Excel

Step 1: Calculate average luminosity.

AWAY is expanding its research on stars. Yellow-white stars have the spectral class “F.” Calculate the average luminosity of all the stars with the spectral classification “F.”

The average luminosity of a yellow-white star is:

The average luminosity for an F class star is 21177.53349

Step 2: Filter the top 5.

AWAY would like a list of the five hottest stars (by temperature) and their color.

STAR	TEMPERATURE	COLOR
1	40,000	Blue
2	39,000	Blue
3	38,940	Blue White
4	38,234	Blue White
5	37,882	Red

The temperature and color of the hottest star is:

> The temperature of the hottest star is 40,000

Part 3 - Gather Insights with Statistics

Tool: Excel Data Analysis ToolPak, Excel functions, and visualizations

Step 1: Calculate and visualize descriptive statistics with the Data Analysis ToolPak.

Report the summary statistics for the *Temperature* column.

> Temperature
>
> Mean 10497.4625
> Standard Error 616.6063847
> Median 5776
> Mode 3600
> Standard Deviation 9552.425037
> Sample Variance 91248824.09
> Kurtosis 0.877352209
> Skewness 1.321568344
> Range 38061
> Minimum 1939
> Maximum 40000
> Sum 2519391
> Count 240

Interpret the skewness and kurtosis for *Temperature*.

The distribution is highly skewed, as the value is higher than 1.

The kurtosis is also highly skewed with a value of 0.877352209

Create and interpret either a histogram or box and whisker plot for *Temperature.* Be sure to paste your visualization below.

Please see excel document for histogram. The majority of temperatures fall in the first frequency. The range is 1,939 to 4,420 and makes scenes due to the Skewness and kurtosis values.

Step 2: Calculate and interpret the correlation of two variables using a scatterplot and the correlation coefficient.

Create and interpret a scatterplot of *Temperature* and *Absolute Magnitude*. Report the correlation coefficient by:

- Displaying the correlation coefficient on the scatterplot
- Using the CORREL function
- Calculating the correlation coefficient in the Data Analysis ToolPak

Be sure to paste your visualization below.

Please see excel file for scatterplot. There is a weak negative correlation between Temperature and Magnitude. The correlation is -0.420260542

Step 3: Use a combination of bar and line charts to compare groups.

AWAY is looking for stars that have similar properties to the sun. Below is the luminosity, radius, and absolute magnitude of the sun:

- Luminosity: 3.75E+28
- Radius: 4.33E+05
- Absolute Magnitude: +4.83

Create a combination chart with *Radius and Sun Radius*. Be sure to paste your visualization below. Note the stars that have a similar radius to the sun.

There are no stars with a radius the same as the Sun Radius in the Data set. I filtered the data for radius between 2.63 and 6 to make the chart. The closest radius is in red on the chat
Please see chart on tab Radius of the workbook.

Create a combination chart with *A_M (absolute magnitude) and A_M Sun*. Be sure to paste your visualization below. Note the stars that have a similar absolute magnitude to the sun.

There are no stars with the same Absolute Magnitude as +4.83 the closest is +5.03 and is in red on the chart. I filtered the chart between 2.41 and 5.5. Please see tab Absolute Magnitude.

Step 4: Create a simple regression equation and interpolate information given new information.

There are many indicators that could predict whether our data contains dwarf stars near plants that might contain life. Choose two:

A. Temperature
B. Luminosity
C. Radius
D. Absolute magnitude

Use the **Data Analysis ToolPak** to create a regression line with the two indicators you have chosen. Use a 95% confidence level. Report your equation below and the value of the correlation coefficient.

AWAY has found a new star with the following characteristics.

- Luminosity: 1.45E+04
- Radius: 3.19E-01
- Absolute Magnitude: -6.12

Use your regression line to find ŷ. What can be said about this new star? Can you predict the color?

> This star would fall below the regression line and to the left. It is must likely that this star would be Super Giant star type.

Part 4 - Plan a Report

Tool: Word document, whiteboard application such as Miro

Step 1: Choose a report style.

Which report style will you use?

A. Annual, quarterly, monthly
B. Compliance
C. Progress
D. Feasibility
E. Operational
F. Strategic
G. Executive
H. Showcase a specific issue
I. Specific sector

Detail why you choose this option.

I choose Showcase a specific issue as the type of report to write because of the information given in the prompt. The question that was given by “AWAY” is “I have been tasked with determining which stars are most likely to have nearby plants that could possibly support life.” Because of this type of specific question, it only makes sense to use this type of report model.

Step 2: Gather report details.

Provide a title for your report based on the main goal or key insight.

The Many Stars in the Sky

Write a brief description (about 2–3 sentences) on what your report is about.

My report looks for correlations between different data types about satellites that might be like our sun and there for have plants orbiting them that might have life on them. The data types that were provided were Temperature, Luminosity, Radis, Absolute magnitude, and Spectral class.

Produce a list of everyone on the team and their roles, such as “created visualizations” or “completed data preparation.”

All parts of this report were created by Brent Zitsman

Step 3: Plan the visualizations.

What will be the main graphic or chart? It should be the most important insight you want to share.

Correlation of Temp & Absolute Magnitude

What will be the supporting graphics or charts? Keep in mind that you might need other visualizations to illustrate the main point and convince your audience.

Are there any other topic-relevant images that you will add to the report for a visual boost? For instance, you might want to include an image of stars or a yellow dwarf in your report.

stars milky way, spaceships the sun

Step 4: Report key insights.

List the main insights you found in your data.

From the data that was provide there is very little correlation between any of them. From the small sample size, it would be very unlikely to find a satellite similar to our sun. It is like looking for a needle in a haystack.

What solution or conclusion will you make? List the insights or data you gathered to support this.

Part 5 - Develop a Data Story

Tool: Word document, whiteboard application such as Miro

Step 1: Complete the data story checklist.

What do you want to do with your dataset?

A. Inform – summarize findings of a study
B. Classify the data
C. Make a company decision or predict future results
D. Inspire/persuade people to act

Who is your audience?

Step 2: Organize your story points.

Choose some common story points for your data story. Write a few details on how you will illustrate these points.

- Change over time
- Relationship of two metrics
- Intersection (when one metric surpasses another)
- Prediction
- Compare and contrast
- Drill down (general → specific)
- Zoom out (specific → general)
- Cluster (values concentrated in an area)
- Outlier (data that lies outside the norm)

Step 3: Create a story arc.

What is the setting (context: who, what, where, when)? Include a hook—something to get the audience's attention.

What are the rising insights that support/lead to your goal or main point?

Step 4: Add context to your story.

Is there any background information the audience needs to know to make sense of the data insights?

Part 6 - Build a Report

Tool: Excel

Create a one-page report (using the ***Part 6_Report Template.xlsx***) that includes:

- Specific, targeted metrics illustrated with meaningful visualizations
- Storytelling techniques
- The recommendation or solution for the client

Consider the following when structuring your report:

- Report goal
- Color scheme
- Visualizations
- Text and graph balance

The final format must be an Excel document that your team will turn in, in addition to this packet.