**| 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?

**The goal of this scenario is to find yellow stars which are most likely to have nearby planets that could possibly support life just like our solar system star, sun, does. The question is that how many yellow dwarf stars are found and are the part of the given database.**

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?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Temperature** | **L** | **Color** | **Spectral\_Class** | **Type** |
| **5800** | **0.81** | **Yellow-White** | **F** | **3** |
| **5936** | **1.357** | **Yellow-White** | **F** | **3** |
| **5587** | **0.819** | **Yellow-White** | **F** | **3** |
| **5300** | **0.59** | **Yellow-White** | **F** | **3** |

**I removed A-M and Radius columns because they are not needed to identify Yellow-White Stars for my goal.**

Step 3: Identify typos.

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

**I cleaned the typos by using consistent format by using color names Yellow-White and Yellow stars.**

**I targeted Yellow-White stars to identify the stars like Sun. Also, we targeted the temperature and the type of the star to make sure it has properties that are like Sun. I used filters**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Temperature** | **L** | **Color** | **Spectral\_Class** | **Type** |
| **5800** | **0.81** | **Yellow-White** | **F** | **3** |
| **5936** | **1.357** | **Yellow-White** | **F** | **3** |
| **5587** | **0.819** | **Yellow-White** | **F** | **3** |
| **5300** | **0.59** | **Yellow-White** | **F** | **3** |

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.

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?

**There were no missing values. So, I selected home, then I selected Find and Select. Then, clicked on special values and clicked on blank; there were no missing/ blank values or cells.**

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.

**There are no duplicate values found.**

# 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:

**Luminosity**

**0.81**

**1.357**

**0.819**

**0.59**

**Average Luminosity**

**0.894**

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 | 40000 | Blue |
| 2 | 39000 | Blue |
| 3 | 38940 | Blue |
| 4 | 38234 | Blue |
| 5 | 37882 | Blue |

The temperature and color of the hottest star is:

**The temperature of hottest star is 40,000, and the color is blue.**

# 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.

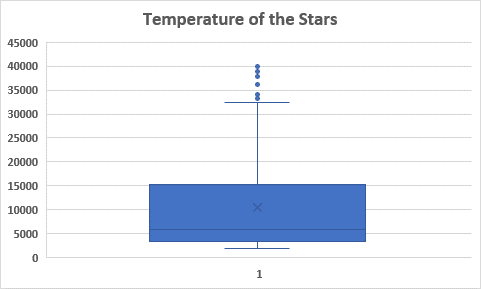
Report the summary statistics for the *Temperature* column.

|  |  |
| --- | --- |
| ***Temperature*** | |
|  |  |
| **Mean** | **10497.46** |
| **Standard Error** | **616.6064** |
| **Median** | **5776** |
| **Mode** | **3600** |
| **Standard Deviation** | **9552.425** |
| **Sample Variance** | **91248824** |
| **Kurtosis** | **0.877352** |
| **Skewness** | **1.321568** |
| **Range** | **38061** |
| **Minimum** | **1939** |
| **Maximum** | **40000** |
| **Sum** | **2519391** |
| **Count** | **240** |

Interpret the skewness and kurtosis for *Temperature*.

**Kurtosis 0.877352209 is the measure of the data if it heavy tailed or light tailed relative to the normal distribution. Positive Kurtosis means that it is heavy tailed. Skewness 1.321568344; Positive value means it is skewed on the right side of the normal distribution curve.**

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



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.

|  |  |  |
| --- | --- | --- |
|  | *Temperature* | *A\_M* |
| Temperature | 1 |  |
| A\_M | -0.42026054 | 1 |

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 is one star which has values close to the sun

|  |  |  |
| --- | --- | --- |
| **Luminosity 0.357** | **Radius**  **1.13** | **A\_M**  **4.78** |

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.

|  |  |  |
| --- | --- | --- |
| **L** | **R** | **A\_M** |
| **0.81** | **0.9** | **5.05** |
| **1.35** | **0.98** | **2.93** |
| **1.357** | **1.106** | **4.46** |
| **0.819** | **0.99** | **5.03** |
| **0.59** | **0.91** | **5.49** |

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 planets that might contain life. Choose two:

1. Temperature
2. Luminosity
3. Radius
4. Absolute magnitude

**I choose Radius and the Temperature to predict whether our data contains dwarf stars near planets that might contain life.**

Use the **Data Analysis Tool-Pak** 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.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.350708 |  |  |  |  |  |  |  |
| R Square | 0.122996 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.118759 |  |  |  |  |  |  |  |
| Standard Error | 8042.261 |  |  |  |  |  |  |  |
| Observations | 209 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 1 | 1.88E+09 | 1.88E+09 | 29.0309 | 1.9293E-07 |  |  |  |
| Residual | 207 | 1.34E+10 | 64677968 |  |  |  |  |  |
| Total | 208 | 1.53E+10 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 7742.662 | 605.4921 | 12.78739 | 5.6E-28 | 6548.9405 | 8936.384 | 6548.94 | 8936.384 |
| 0.17 | 6.021571 | 1.117583 | 5.388032 | 1.9E-07 | 3.81826718 | 8.224874 | 3.818267 | 8.224874 |

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?

|  |  |  |  |
| --- | --- | --- | --- |
| L | R | A\_M | Color |
| **1.43** | **0.99** | **2.41** | **yellow-white** |
| 1.35 | 19 | 2.93 | yellow-white |
| 1.357 | 23 | 4.46 | yellow-white |

**The new star will be yellow-white based luminosity.**

# 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?

1. Annual, quarterly, monthly
2. Compliance
3. Progress
4. Feasibility
5. Operational
6. Strategic
7. Executive
8. Showcase a specific issue
9. Specific sector

Detail why you choose this option.

**Report type will the “Showcase a specific issue.”**

Step 2: Gather report details.

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

**Alien’s Yellow Star**

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

**The report of this scenario is to find yellow stars which are most likely to have nearby planets that could possibly support life just like our solar system star, sun, does. The question is that how many yellow dwarf stars are found and are the part of the given database.**

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

Step 3: Plan the visualizations.

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

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.

Since I am trying to locate the star which has relatively close values of Luminosity, temperature, size, and type to the Sun so it can be suitable for the survival of life, I will use the filters and will create the table to narrow down the star search close to the Sun, and compare it with the Hertzsprung-Russell Diagram below to show the relevance. Bolded values are closer to the Sun, and the values can be eyeballed for the Hertzsprung-Russell Diagram given below.

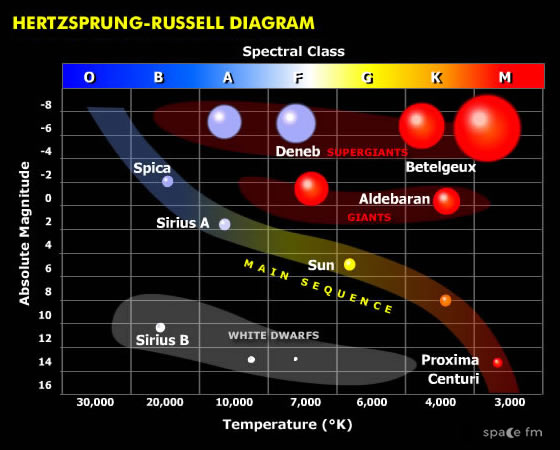
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Temperature | L | R | A\_M | Color | Spectral\_Class | Type |
| 6757 | 1.43 | 1.12 | 2.41 | yellow-white | F | 3 |
| 6380 | 1.35 | 0.98 | 2.93 | yellow-white | F | 3 |
| **5936** | **1.357** | **1.106** | **4.46** | **yellow-white** | **F** | **3** |

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.

A picture containing blur, night sky

Description automatically generated

Figure : Yellow-White Dwarf Star



Step 4: Report key insights.

List the main insights you found in your data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Temperature** | **L** | **R** | **A\_M** | **Color** | **Spectral\_Class** | **Type** |
| **5936** | **1.357** | **1.106** | **4.46** | **yellow-white** | **F** | **3** |

**The temperature of the star in the table is almost like the Sun and it is 5936 K, and temperature of the Sun is 5778 K to 6000 K according to NASA website. Luminosity of the Sun is 3.28 and the star I targeted has 1.35. Besides, the absolute magnitude of the Sun is 4.83 and the star I targeted has 4.46. The color of the spectral class is G for the Sun and F for the star. Sun is type 2 star and the targeted star is type 3.**

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

**I was able to narrow down to the star which has temperature, absolute magnitude close to the sun, and there are differences for the values of the radius, luminosity and spectral class. I am positive that may be there is life for the aliens on the near by planets of the star; however, the alien might have variation in the required amount of light and energy.**

**Recommendations for the database is to include comparison with famous stars especially the Sun because I had problem finding the correct values and units for the luminosity/ temperature/ radius/ and star type. There must be no typos in nomenclature of the colors. It will be useful to name the stars and their properties in correct units for easy conversions, comparison and contrast of two or more than two stars.**

# 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?

1. Inform – summarize findings of a study
2. Classify the data
3. Make a company decision or predict future results
4. Inspire/persuade people to act Who is your audience?

**Inform\_ summarize the findings of a study.**

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)

**Compare and contrast/ Drill down (general to specific)/ Prediction**

Step 3: Create a story arc.

**Setting: Jane a space scientist is working in the lab and receiving signals from the close by galaxy in the space and feeling excited to receive some radio signals which she never received. With all excitement she is thinking if there is a planet which can support the alien life. She pulls out the database “Stars” from the company AWAY; she cannot access it because it is restricted. This 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. Jane contacts the company and tells them what happened, and they ask her to work for them and share her findings. Now she is trying to find out which stars are most likely to have nearby planets that could possibly support life.**

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

**Rising insights: Jane the space scientist cannot access the database “Stars” without joining the company AWAY. She joins the company to drill down the data to predict which stars possibly can support the forms of life or aliens. Her goal is to find out the star which has properties close to the Sun at least temperature, color, and absolute magnitude wise. Her goal is to find yellow stars which are most likely to have nearby planets that could possibly support life just like our solar system star, sun, does which most likely is the planet of the star where she is receiving signals from. The question is that how many yellow dwarf stars are found and are the part of the given database.**

Step 4: Add context to your story.

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

**Context and the background of the story: Jane is already researching on receiving signals from space from aliens who are situated in the close by galaxy and the solar system like humans. She receives the signals and the reply of her signals. She received signals twice in the result of her sent radio signals to the space. She is hopeful that she can narrow down her search for the specific star which she is receiving signals from.**

# 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.