**Analysis of Olympics Data**

This audit trail has been produced by A. N. Actuary on 10th October 2021. It relates to the spreadsheet ‘Module 5\_Olympics Analysis.xlsx’.

**Objective**

This spreadsheet has been produced for the government of Country A. The purpose of the spreadsheet is to analyse historic data related to the men’s and women’s Olympic high jump and long jump events to help the Government determine their Olympic funding plan. Data has been provided for historic medal winners across the period 1980-2020, and the winning distances across the period 1964-2020. The Government has also informed us that they have $50m to invest in athlete development, and the following options are being considered when determining how to split the funding across the four events:

*Option 1:* Allocate twice as much money to high jump events as to long jump events, and allocate 60% of the total funding to women’s events.

*Option 2:* Allocate the funding based on the factor multiples set out in the following table, where $x$ is an amount in $ that will depend on the total funding amount and total number of medals.

|  |  |
| --- | --- |
| Number of medals | Type of event |
| Men’s | Women’s |
| 0 | 4$x$ | 6$x$ |
| 1 | 3$x$ | 4.5$x$ |
| 2 | 2$x$ | 3$x$ |
| 3 | $$x$$ | 1.5$x$ |

The spreadsheet calculates the number of medals that the Country B, Country A, and Country A’s main competitors have won over the period 1980-2020, and shows Country B’ medals in a chart. It also calculates the average increase in winning distance for each event, and shows the historic winning distances on a chart. Finally it calculates how the total funding should be split based on the two options being considered.

**Data and Data Checks**

The government of Country A has provided two sets of data, and these are shown on the sheet ‘Data’. The first set shows the historic medal winners across each event for the period 1980-2020. The data shows the sex and team of each medal winner, and the year, event and type of medal (bronze, silver or gold). The second set shows the winning distance in metres for each event for the period 1964-2020. The data has been copied onto the sheet ‘Data checks’.

The following validation checks have been performed on the data:

* The check in column H checks that the year for the first data set is in the range 1980-2020.
* The check in column I checks that each ID number is 1 greater than the previous ID number.
* The checks in cells M7 and M8 check that half of the data points are male and half are female.
* The checks in cells M10 to M12 check that there is an equal number of gold, silver and bronze medals.
* Auto filter has been used to check that for all men’s events sex = M and for all women’s events sex = F.
* Visual inspection of the data shows that there are no blank values. This is also checked in row 138.
* The check in column U checks that the year for the second data set is in the range 1964-2020.
* The checks in columns W to Z check that the men’s distances are longer than the women’s (for the same event) and that the long jump distances are longer than high jump (for the same sex).
* The check in row 21 checks that there are no blank cells.
* The maximum, minimum and averages distances have been calculated for each event. These look reasonable and there are no obvious outliers.
* Visual inspection of the data shows that there is only one entry for each year.

The checks all give satisfactory results and so the data is assumed to be accurate.

**Assumptions**

The following assumptions have been made in the model:

* The data provided is complete and accurate.
* The Olympics are held every four years.
* In each event there is a single medal winner for each of the bronze, silver and gold medals.
* No adjustment is made to allow for the fact that the 2020 Olympics was actually held in 2021.
* Long jump distances are generally greater than high jump distances, and men’s distances are generally greater than women’s.
* 100% of the $50m of funding will be allocated across the four events.
* Only males can compete in men’s events, and only females can compete in women’s events.
* The 2024 Olympics will go ahead and will include the high jump and long jump events.

**Parameters**

This sheet sets out the parameters used in the model:

* The total funding available is $50m.
* The team whose results are shown in a chart is Country B.
* The teams whose results are shown in a table are Country A, Country C, Country D, Country E and Country F.
* The high jump to long jump ratio for Option 1 is 2.
* The women’s funding percentage for Option 1 is 60%.
* The factors for Option 2 are as shown in the table in the ‘Objective’ section above.

Each parameter has been set as a named range, with the name shown in grey next to the parameter. These parameters are used in the other sheets in the model.

Note that there is also a user input on the sheet ‘Funding calculations‘. This is explained further below. All cells with user inputs are coloured blue.

This sheet also shows the colour coding used in the model.

**Medal Analysis**

This sheet first calculates the total gold, silver and bronze medals for Country B using the COUNTIFS function. The total number of medals is simply the sum of these three values. This data is also shown in a bar chart.

There is then a table repeating this calculation for Country A and its main competitors.

Note that both calculations link to the teams entered on the ‘Parameters’ sheet.

The following checks have been performed:

* Cell G6 checks that the total number of medals for Country B (Gold plus Silver plus Bronze) matches the total calculated directly from the data using a COUNTIF formula.
* Cells I22 to I26 repeat this check for the other countries.

**Distance Analysis**

This sheet shows the historic data for the winning distance for each of the four events (linked to the ‘Data’ sheet). Cell C22 calculates the total number of Olympic periods. In row 23 the total percentage increase in distance is calculated as: (2020 distance/1964 distance) – 1.

In row 24 the percentage increase per Olympics is calculated as: (1 + total increase) ^ (1/number of periods) – 1.

In row 28 the expected 2024 distance is calculated as: 2020 distance \* (1 + increase per Olympics).

The historic winning distances for each of the four events are shown in a line chart.

The following checks have been performed:

* The checks in row 26 check that applying the percentage increase to the 1964 distances replicates the actual distances in 2020 for each of the four events.
* The check in row 29 checks that the 2024 expected distance is greater than the 2020 distance for each event (as the average increase is positive for all events).
* Visual inspection of the chart shows that all of the lines trend slightly upwards over time, which is consistent with the positive change in distances calculated.

**Funding Calculations**

This sheet determines how the total funding should be split between the four events using the two options set out in the ‘Objectives’ section above.

Cell C2 is a user input and it is currently set to Country A. Cells D5 to D8 calculates the medals for each event for the team entered using the COUNTIFS function.

The calculations for Option 1 use the inputs from the ‘Parameters’ sheet. The total funding for women’s events is calculated as: total funding \* women’s funding %. The total funding for men’s events is calculated as: total funding – total funding for women’s events.

For each sex the funding for high jump is calculated as: total funding for sex \* 2/3. The funding for long jump is calculated as: total funding for sex – funding for high jump.

The calculations for Option 2 use INDEX and MATCH functions to determine which factor multiples apply based on the total number of medals for each event. Cell D33 calculates ‘x’ as: total funding/total factor multiples.

Cells D36 to D39 then calculate the funding for each event as: factor \* event factor multiple.

The following checks have been performed:

* Cell F8 checks that the total number of medals across the four events matches the total across gold, silver and bronze calculated on the sheet ‘Medal analysis’.
* Cells F19 and F21 check that twice as much funding is allocated to high jump as to long jump.
* Cells F23 and F40 check that the total funding across the four events sums to $50m.