**Titanic Dataset Power BI Exercises**

**Column Headers and Column Information:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Data Type** | **Valid Values (Range)** | **Explanation** |
| PassengerId | Whole number | 1 - 891 | Unique integer passenger Id |
| Survived | Whole number | 0, 1 | 0 - Died, 1 - Survived |
| Pclass | Whole number | 1, 2, 3 | Passenger Class |
| Name | Text | Any | Note consistent format: <Last name>, <Title>. <First name> (with extra rules for married women) |
| Gender | Text | male, female | Gender |
| Age | Decimal number | <100 |  |
| SibSp | Whole number | Any | Number of siblings (brothers/sisters) and spouses in the party travelling with the passenger |
| Parch | Whole number | Any | Number of parents and children in the party travelling with the passenger |
| Ticket | Text | Any |  |
| Fare | Decimal number | > 0 | In GBP (1912 value) |
| Cabin | Text | Any |  |
| Embarked | Text | S, Q, C | Point of embarkation: S - Southampton, Q - Queenstown, C - Cherbourg |
| **Note that some columns e.g. Age, Embarked have missing values** | | | |

**Data Cleaning Steps**

**Import dataset**

**Select Passengers Table**

**Transform data using Power Query Editor**

1. **Check Data Types**
   * Ensure datatypes for each column match the contents of cells.
2. **Replace Values in 'Survived'**
   * Check that datatype is appropriate. If not text datatype, change the datatype to text.
   * Use Replace Values to replace 0 in the Survived column with "Deceased" and 1 with "Survived". Rename column to "Survival".
3. **Replace Values in 'Pclass'**
   * Check that datatype is appropriate. If not text datatype, change the datatype to text.
   * Use Replace Values to replace 1 in the Pclass column with "1st", 2 with "2nd", and 3 with "3rd". Rename column to "Passenger Class".
4. **Create Family Size Column**
   * Use Add Column tab and Custom Column to create a new Family Size column that is SibSp + Parch + 1 (adding one to include the passenger themselves). Check datatypes and set to whole number.

A screenshot of a computer

Description automatically generated

1. **Replace Values in 'Embarked'**
   * Use Replace Values to replace the initials of the location: C with "Cherbourg", S with "Southampton", and Q with "Queenstown". Check "match entire cell contents" in advanced options of the "replace values" dialog box.
2. **Split Columns**
   * Use Split Column under Transform tab twice to create a Surname column and a Title column. Rename these columns as needed.
   * A screenshot of a computer

     Description automatically generatedUse format to trim away spaces in the title column. (You can find it under Transform tab and Format dropdown.)
3. **Remove Columns**
   * Remove Cabin column from dataset.
4. Handling Missing Values.
   * Filter null values on Embarked column. (deselect null values)

Select Close&Apply to return report view.

**Report Creation Exercises**

1. **Create a Visual for Passenger Demographics**
   * Include a pie chart for gender distribution. (Use Gender and PassangerID Fields)
   * Include a bar chart for class distribution. (Use Passenger Class and PassangerID Fields)
2. **Create a Visual for Survival Analysis**
   * Include a bar chart for survival rate. (Use Survived and PassengerID Fields)
3. **Create a Visual for Embarkation Points**
   * Create a treemap visual to analyse the number of passengers and their embarkation point.
   * Use filter to show only survived passengers.

**Additional Visualization Exercises**

* + Create another visual of your choice.

**Use learned skills to colour, format, filter and edit your visuals.**

**Add report title**

**Publishing and Sharing in Power BI Service**

1. **Publish to Power BI Service**
   * Once your report is complete, publish it to the Power BI Service.
     + In Power BI Desktop, click on the "Publish" button in the Home tab.
     + Sign in with your Power BI account using bootcamp email.
     + Select the destination workspace where you want to publish your report.
     + Wait for the publishing process to complete and take note of the link to the published report.
     + If required sign into power bi service using bootcamp email.
     + Click on share button at the top of your report and search and select [your](mailto:yusufs@justit.co.uk) trainer.
     + Click on send to share your work with the trainer.