We first need to prepare the data for better readability and understanding, I first found blank values in the whole dataset by selecting the entire sheet and using **go to special** function and selected blanks and filled them with red colour and marked them as “Not Available” for better understanding and easy traceability of all the missing values. Imputations are used to fill these missing values with data wherever possible.

Created a Macro named “Preparing\_the\_Heading” that will automatically add background colour to the headings, auto fit column width for all records, freeze the top row using freeze pane and will also display the heading in bold. To use this macro, select the entire data and press “Ctrl+Shift+H”.

**First Name**

* Used the function **Proper** in a macro that can be accessed using “Ctrl+Shift+P” which will automatically capitalise the first letter of a word.
* Assigned the first blank cell with “V” based on the email ID.

**Last Name**

* Used the macro created earlier using the function Proper to have the first letter of the word capitalised.
* Removed special characters “Ã¸” from “JÃ¸rgensen” and converted the name to “Jensen” based on the email ID.

**Email**

* Created a macro that uses the function **lower** to change the entire text into lowercase as email IDs are not case sensitive and the email servers will read the email address the same way as long as the letter and numbers (if any) match the official email address.
* “ncamp@metro” had a **dot(.) com** missing
* marla.biesche@gm com had a **dot(.)** missing after gm.

**Job Title**

* This column had some blanks that were filled with text “Not Available” so that there are no blank cells.
* Changed “Sr Mgr” to “Senior Manager”.
* Changed “Mktg Strgist” to “Marketing Strategist”.
* Created a macro that uses the trim function to remove extra spaces between text.

**Company Name**

* Added spaces between “WarnerBrosUK” to make it “Warner Bros. UK.
* Removed special characters “Ã©” to have the correct name “Condé Nast”.

**Associated company**

No errors were found in the Associated Company column.

**Phone number**

Phone numbers had values missing completely at random, we cannot impute them with anything and have filled it the with “Not Available” so that there are no blank cells.

**Street Address**

* Changed the heading “Street Address” to “Address”.
* We can use the internet to find the address for companies that have incomplete address as done for Company name Georg Jensen, Dirty loft and Abril.

**Company Size**

* Plotter a scatter plot to get a visual of outliers.
* then, removed all the outliers using the Inter Quartile Range method where we first calculate the first quartile then the third quartile and subtracted the 3rd Quartile from the 1st Quartile to find the Inter Quartile Range(IQR), after removing the outliers we then calculated the mean as 1,113.56 and imputed the rounded off value (1,114) in all the blank cells under company size.

Formula for IQR= Q3-Q1

Formula for calculating upper and lower range:

Lower Range = Q1-(1.5\*IQR)

Upper Range = Q3+(1.5\*IQR)

Any value Above the Upper Range and Below the Lower Range are considered to be outliers.

* Changed the format from general to number and added comma (,) for better readability.

**Contact Owner**

* Took the mode using mode function and sum ifs of all the names available to find that that the name ”Katy” had the highest mode, we have put “Katy” in th two columns that had blank values and have kept the unassigned values the same because there might be chance that the company may not have provided us with a contact name, but no cell in this column is empty.

**Annual Revenue**

* There were 21 columns with missing values, we can impute these values with the mean or median if the data set is large enough since the data set is of a smaller size imputing with mean or median would make the variance and standard deviation to reduce significantly which could result in faulty/incorrect analysis.