As a way to showcase my learned skills in Excel, SQL and Tableau I decided to take on two datasets and develop them into dashboards. In this way I demonstrate my ability of having learned how to clean data, how to generate outputs and how to visualize these outputs in three of the most common software used for data analysis.  
I have decided to learn cleaning data with Excel, learn how to manipulate data to get outputs with SQL and learn how to visualize data with Tableau.  
  
The two datasets I picked are sample datasets provided by Kaggle and Tableau. These are datasets available to the public for everyone to practice with.  
I picked one dataset under the business section and one dataset under the public data section. The datasets are called ‘Superstore Sales’ and ‘New York City Airbnb Open Data’. The Superstore sales dataset concerns a fictitious company in the United States. This dataset for example has information regarding sales, products, returns and profits.  
The Airbnb dataset contains Airbnb listings in New York City and information regarding the boroughs and neighborhoods the listings are in. well as prices and the type of rooms they are.  
  
I found these to be satisfactory choices because they are both realistic scenarios where I could be employed in. For example in the future in my career I could get stores as clients who want me to perform an analysis on their performances. Or I could even find myself working in the financial department of a store.  
Same applies for the Airbnb listings. I also have the ambition of visiting New York City some day and found this dataset quite intriguing to perform data analysis on.  
  
To demonstrate my skills I cleaned up both datasets with Excel. Both sets were visualized with Tableau as well. I performed SQL queries on the superstore dataset to generate certain outputs someone in a business setting might want to see.  
  
**Airbnb dataset**  
I started with exploring the dataset and taking note of the different columns. The first step in data cleaning I took was to remove entries with empty inputs. I applied conditional formatting to the dataset and let empty cells be colored in yellow. Then by filtering on color I was able to spot the empty cells and deleted the entries with the empty cells.  
  
Next I noticed the inputs for latitude and longitude varied. For example one entry could have a latitude of 40,589 while another could have it for 4,0656. By using the LEFT and RIGHT formula in Excel I was able to strip the first two numbers from the rest and insert a comma in between. This allowed the latitude and longitude (for longitude it was the first 3 inputs separated) to be transformed in the correct input in a matter of seconds.  
  
Furthermore there was a column with information about the amount of reviews there were per month. For example an entry could be 2 but also 2.5. Depending on what part of the world you open the Excel sheet in, 2.5 will not be seen as an integer as it was in my case. I used Excel’s search and replace function for this column to replace every dot with a comma, thus turning it into an integer.  
  
After that I formatted cells into their correct datatype. For example making sure the ID columns are converted to integers. Text columns into text format. The price in dollars. There were columns concerning dates, which I converted into the date datatype.

To finish it off I looked into duplicate entries and anomalies. I searched for double entries by using the TEXT.JOIN formula in Excel. This basically combines an entry into one string. So by combining columns into one string and then applying conditional formatting to them by coloring duplicated entries in a certain color you are able to filter them out. Next I looked for anomalies in prices. A price of zero dollars is obviously not correct and should be removed from the dataset. I looked for the dates of reviews, I noticed a review from 1900 which was probably some error but still had to be removed. I checked the boroughs, the dataset was about New York City so it should have only included listings in New York City.  
Counting from the first column down excluding the header there were 49.074 entries. After cleaning up the dataset I ended up with 38.530 entries excluding the header.  
  
The next step for me was to visualize this dataset. I chose to show a map of New York City with the average prices per neighborhood displayed on them. The highest price was the darkest color and so it faded with each lower amount. To get this view I put the longitude in the columns and the latitude in the rows. I put the neighborhood and borough columns as marks in the software and that produced the output as can be seen in my dashboard. I also wanted to show the average price per neighborhood in a horizontal bar for easy readability because the map can give you a quick glance of the hotspots without showing prices right away while the horizontal bar can give a quick breakdown of which specific neighborhood has which average price.  
Other visualizations I have opted for were a donut chart of the representation of listings per borough compared to the total listings, the most booked types of rooms per borough and the reviews per month.

I made sure to make everything interactable and interactive, meaning if you were to click on a certain borough on the donut chart it will for instance showcase the average prices on the map for only that specific borough. Every visualization will automatically adjust to that borough. And this can be done with any kind of metric in the dashboard.  
  
With the dashboard I made users can get certain information very easily.  
For instance, users can learn what percentage of total listings are located in which borough. This can help in making decisions of in which borough to invest and see which borough could perhaps have less competition.   
They can learn what type of rooms are booked the most in which boroughs. This can help in deciding what type of Airbnb to start themselves.   
Users can see which neighborhood has which average prices, this can help in deciding where to start your Airbnb or even for non-financial users this can help you in picking a neighborhood that fits within your budget for your vacation to New York City.   
Furthermore, I have shown the reviews per month, this helps to see when the city has its busiest periods. It can help in deciding when to book an Airbnb and it can help with hosts deciding to lower their prices in the less busier months as a strategy to attract more people in those months.  
<https://public.tableau.com/app/profile/burak.gunaydin/viz/DatavisualizationairbnbNYC/Dashboard1>  
  
 **Superstore sales dataset**  
This dataset did not have empty entries as it seemed already to have been cleaned but I still explored it to make sure every column had the correct datatype and that there were no strange entries. The dataset had 3 tables. One table had the orders that were returned. The second table had people listed who are presumably the regional managers. And the third table had the performance of the company in it.  
  
For these datasets I performed SQL on them to get certain outputs that would be necessary in a financial environment. I imported the datasets into MySQL.  
An example of a wanted output in here could have been that the managers wanted to know what the profit is per state.  
The following query gives us that output:  
SELECT State, SUM(Profit) AS Total\_Profit\_Per\_State  
FROM superstore.performance  
GROUP BY State;  
With this query we are telling it to calculate all of the profits per state from the performed table and then group it per state.   
  
Another wanted output could have been wanting to know how much the categories are selling for and what percentage of total sales they are.  
SELECT Category, SUM(Profit) AS Total\_sales\_Per\_Category,  
CONCAT(ROUND((SUM(Sales) / (SELECT SUM(Sales) FROM superstore.performance)) \* 100, 2), '%') AS Sales\_of\_total   
FROM superstore.performance  
GROUP BY Category;

If we want to apply the same for profit it is quite similar:  
SELECT Category, SUM(Profit) AS Total\_Profit\_Per\_Category,  
CONCAT(ROUND((SUM(Profit) / (SELECT SUM(Profit) FROM superstore.performance)) \* 100, 2), '%') AS Profit\_of\_total   
FROM superstore.performance  
GROUP BY Category;

If we wanted to go more in depth we could ask for an output of the best-selling sub-categories since they will give more information.  
SELECT `Sub-Category`, SUM(Sales) AS Total\_Sales\_Per\_Category,  
CONCAT(ROUND((SUM(Sales) / (SELECT SUM(Sales) FROM superstore.performance)) \* 100, 2), '%') AS Sales\_of\_total   
FROM superstore.performance  
GROUP BY `Sub-Category`;  
Notice that Sub-Category is between ``, that is necessary because of the dash. `` makes it understand that we are looking for a column named Sub-Category, because the dash is a different function in SQL so it would not have given us the output we were looking for.  
  
Another realistic output is wanting to know what the returned orders are per category and per state:  
SELECT p.State, p.Category, COUNT(\*) AS Returned\_Orders\_Count  
FROM superstore.performance p  
JOIN superstore.`returns` r ON p.`Order ID` = r.`Order ID`  
WHERE r.Returned = 'Yes'  
GROUP BY p.State, p.Category  
ORDER BY p.State, p.Category;  
  
The same can be done for the sub-categories by replacing p.Category with p.`Sub-Category`.

Another important output is wanting to know what the sales are per customer segment per state. There are three types of customers: consumer, corporate and home office.   
SELECT State, Segment, SUM(Sales) AS Total\_Sales  
FROM superstore.performance  
GROUP BY State, Segment;  
  
You could go more in depth and want to know what sub-category is most popular within segments per state.  
SELECT State, Segment, `Sub-category` AS Best\_Selling\_Subcategory, SUM(Sales) AS Total\_Sales  
FROM superstore.performance  
GROUP BY State, Segment, `Sub-category`  
ORDER BY State, Segment, Total\_Sales DESC;  
  
These have all been examples with SQL in acquiring outputs that can help managers make certain decisions.  
Some questions that can be answered with the queries we have done here are:  
- What are the earned profits per state? And what states have the highest profits? Which ones have the lowest?  
- What is the percentual category sales distribution compared to total sales? What are our most selling categories? What are our most profitable categories?  
- What is the percentual sub-category sales distribution compared to total sales? What are our most selling sub-categories? What are our most profitable sub-categories?  
- Which categories are returned the most? And which states have the most returns?  
- Which sub-categories are returned the most? And which states have the most returns?  
- What is the sales distribution per customer segment per state?  
- What customer base buys which sub-categories the most in which state?  
  
Just with these queries alone a manager could have lots of outputs for analysis. I want to show here that once you master SQL your outputs can be limitless, it is essentially like doing a Google search within a database.

The next step is to visualize some of these outputs in a dashboard.  
The queries are good for managers who want to analyze data but a dashboard is better for presentations so people can quickly absorb the information without reading a bunch of lists.  
The dashboard shows a map of the United States with each state showing the corresponding profit. The states that are running on losses are red as it gradually goes to blue for the states that make a profit. This shows in one quick glance which states are doing well and which ones are not doing well.  
  
Next there is a top selling sub-categories bar chart. A chart of returns per category per state. And line charts which show sales per category per quarter and sales per segment per quarter. The dashboard is interactive and interactable as well because you can click on any point of any chart and the rest of the dashboard will adjust itself to the information you want to see. This is a helpful dashboard because it can show you the profit per state, top selling sub-categories, returns per category per state, sales per category per quarter and sales per segment per quarter.  
With this dashboard the user can quickly gain knowledge on which states are doing well and not so well so that they can adjust their strategy per state. They can see which states have the most returns, this allows them to research why those states have those kinds of returns. They can see what the sales are per segment and how this has developed over time. And they can see what the sales are per category and how this has develop over time.  
<https://public.tableau.com/app/profile/burak.gunaydin/viz/PerformancevisualizationSuperstoreU_S_/Superstoredashboard>