جامعة الإسكندرية
ALEXANDRIA UNIVERSITY

Faculty of Computers and Data Science

Intelligent Systems Department

Field Training Project

Analysis and Report Design

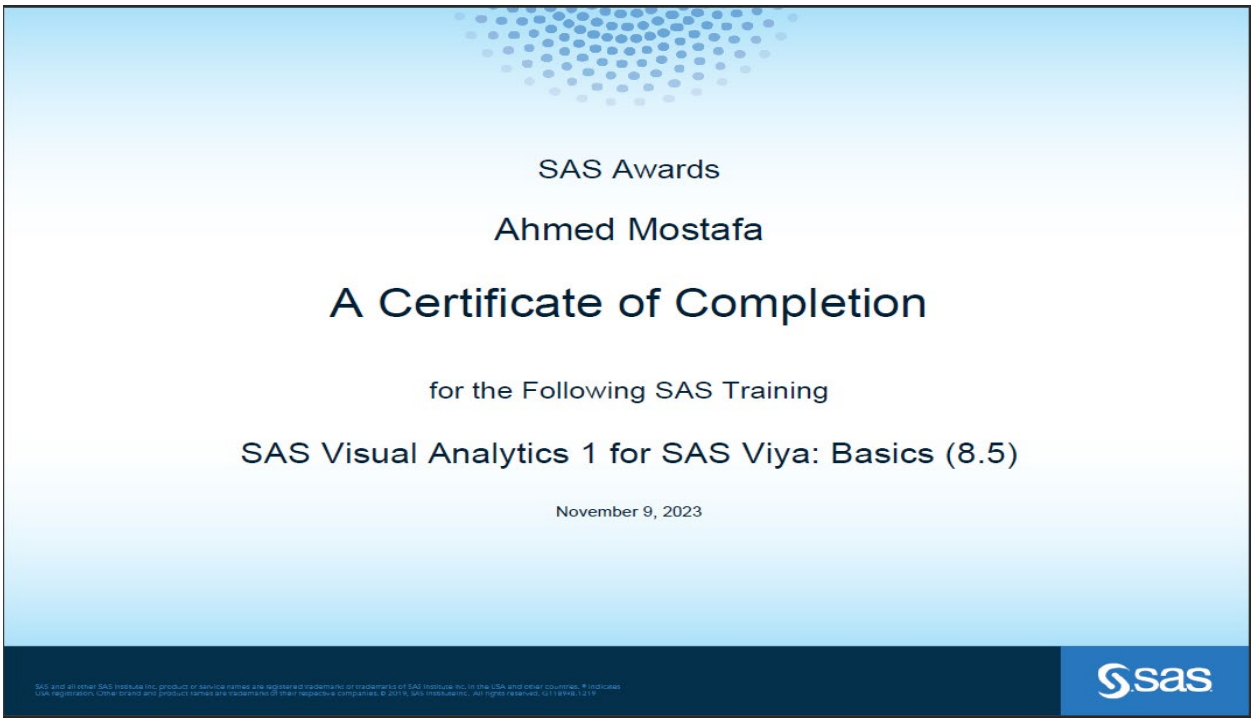
Team Members

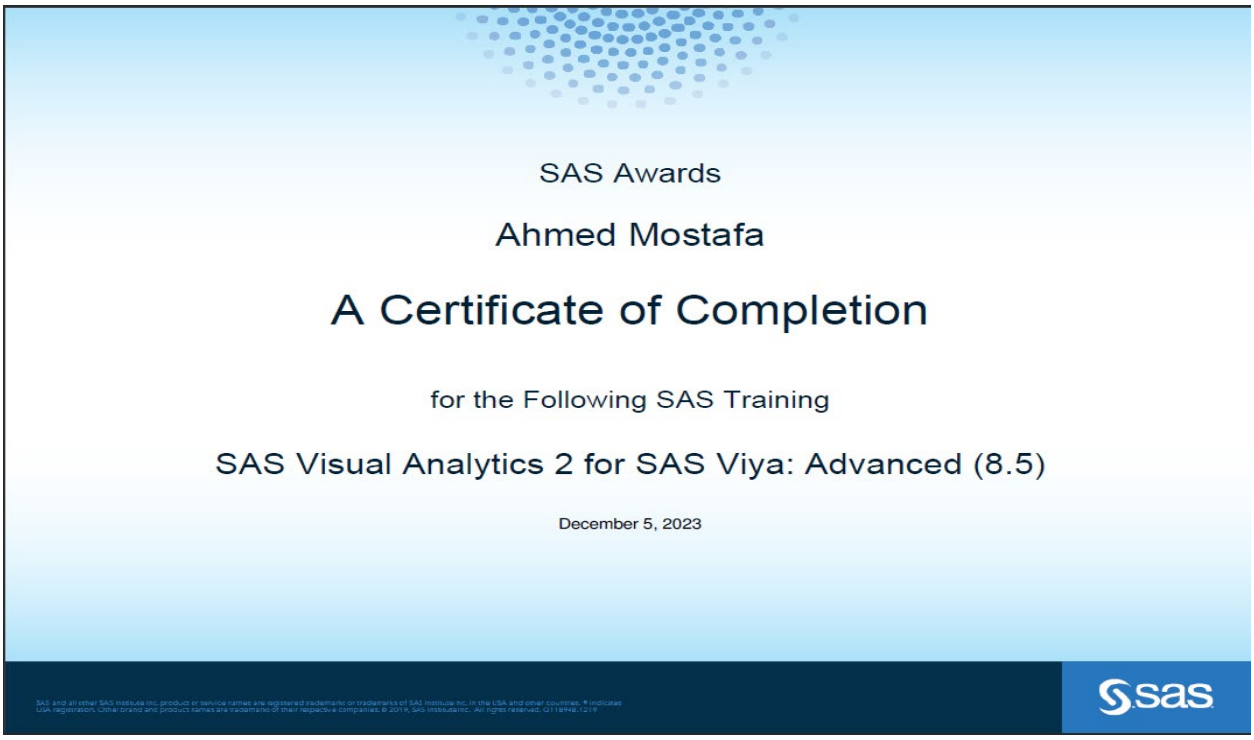
Name	ID	T*asks*
Ahmed Mostafa AbdelRahman	**20221372883**	• **Access** • **Investigate**
Abdelrhman Mohamed Abdelhady Hodib	**2022513643**	• **Prepare** • **GARAGE AREA AND LOT AREA ON HOUSE PRICE** • **Report**
Ahmed Ehab El-Sayed	**20221445838**	• **HOME AGE AND SALEPRICE** • **HEATING QC AND CENTRAL AIR ON SALE PRICE**
Mazen Gaber Mahmoud	**20221372110**	• **GROUND LIVING AREA AND SALEPRICE** • **SEASONS AND SALEPRICE**

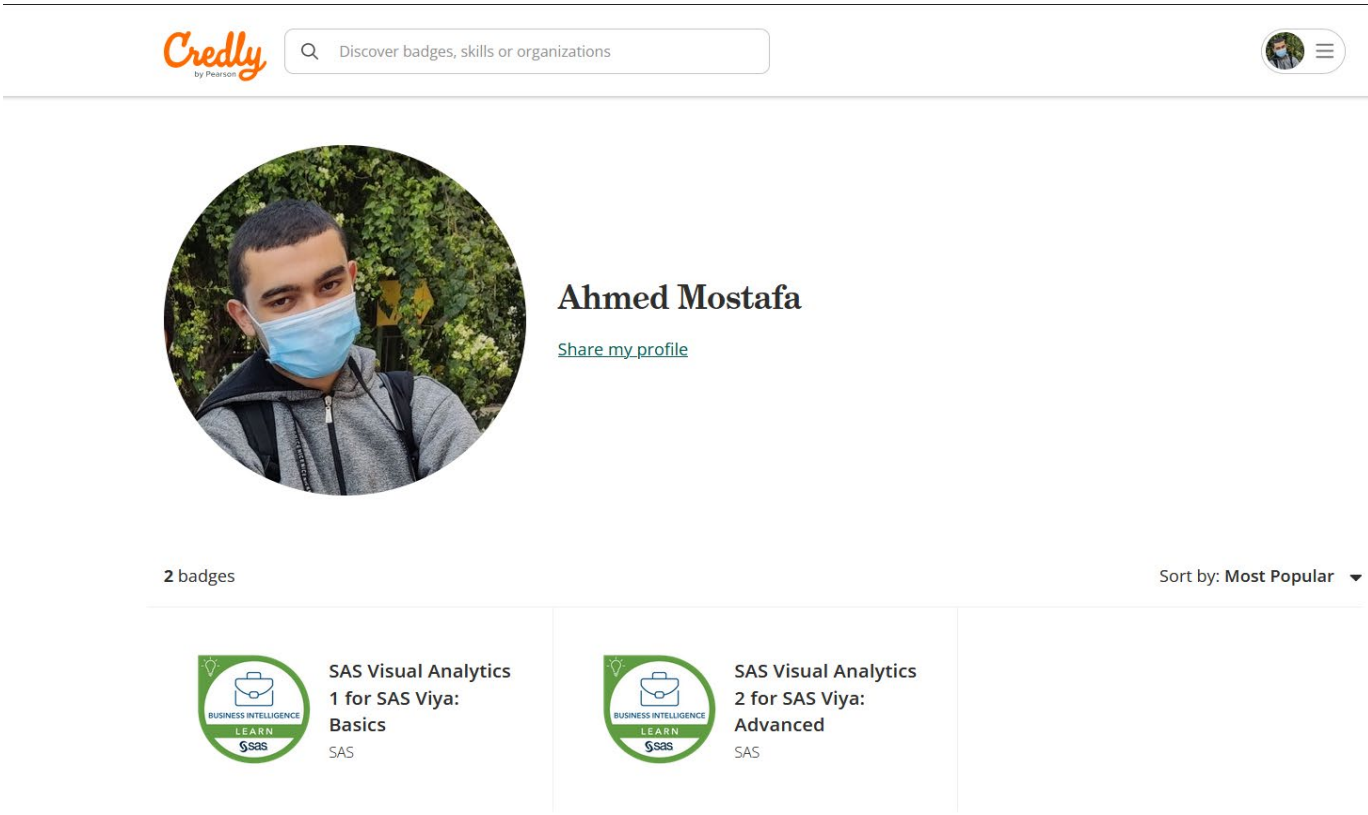
Contents

1 Certificates

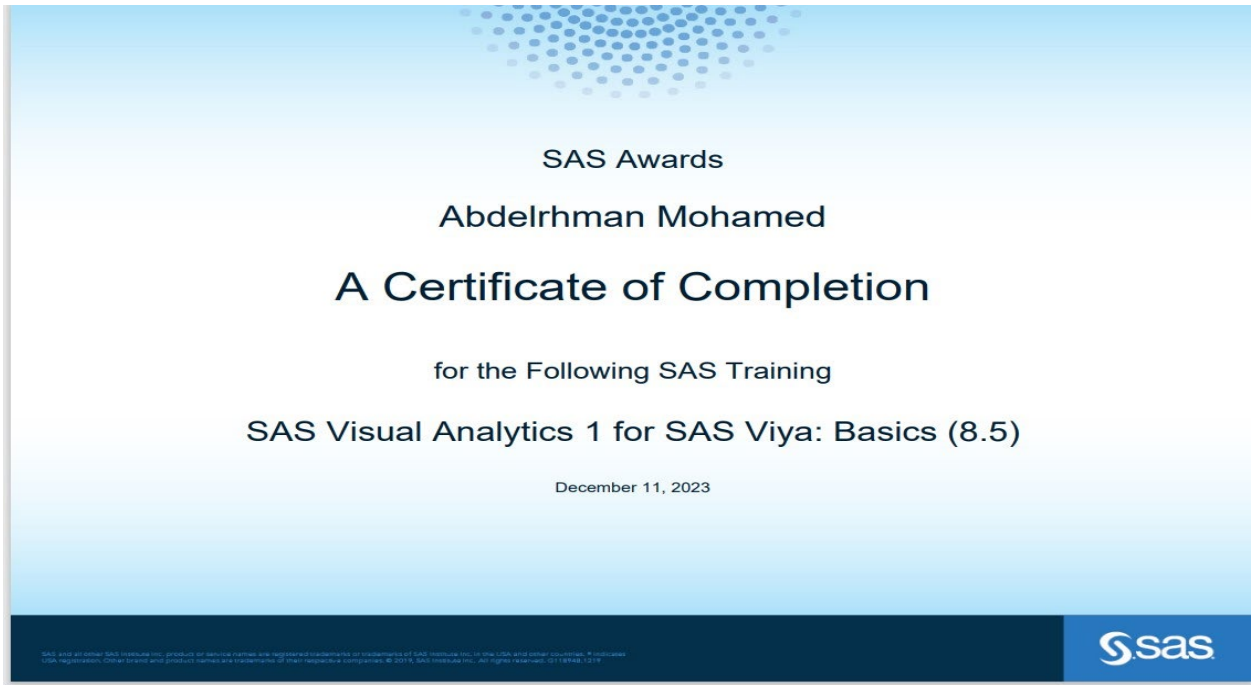
1.1 Ahmed Mostafa AbdEl-Rahman 20221372883

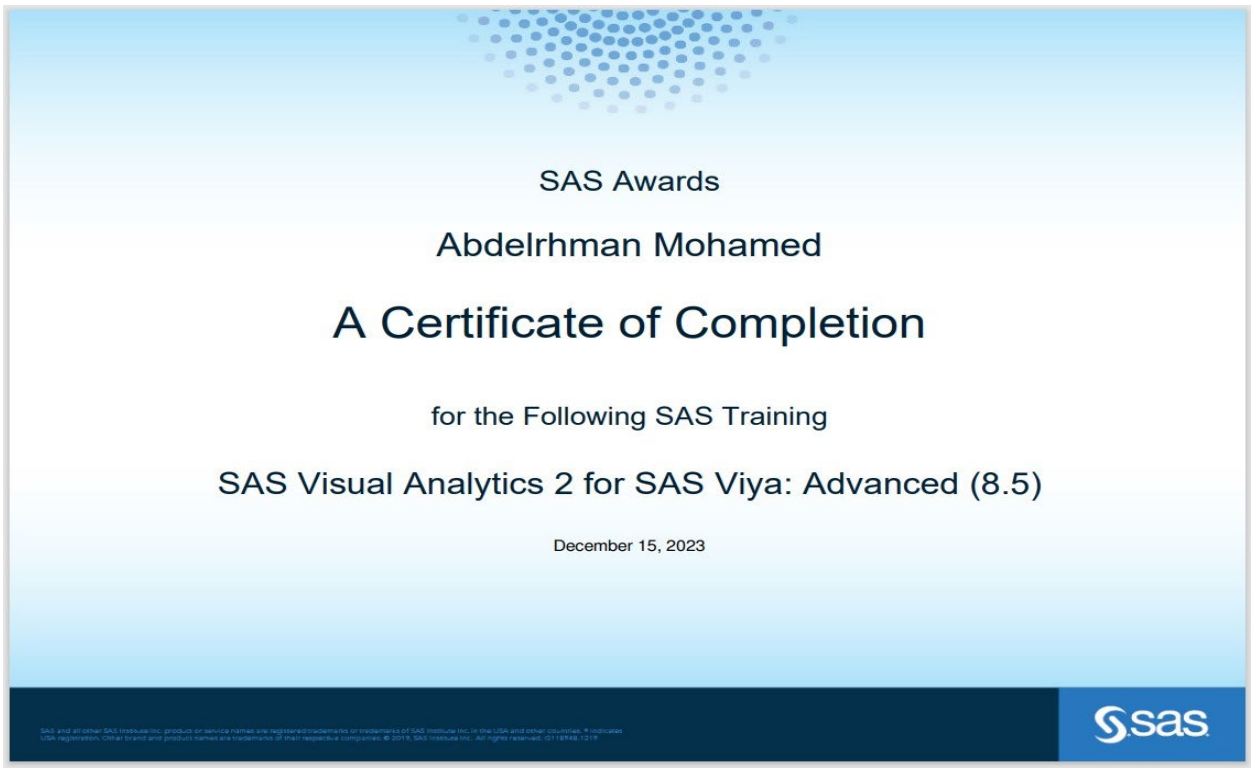


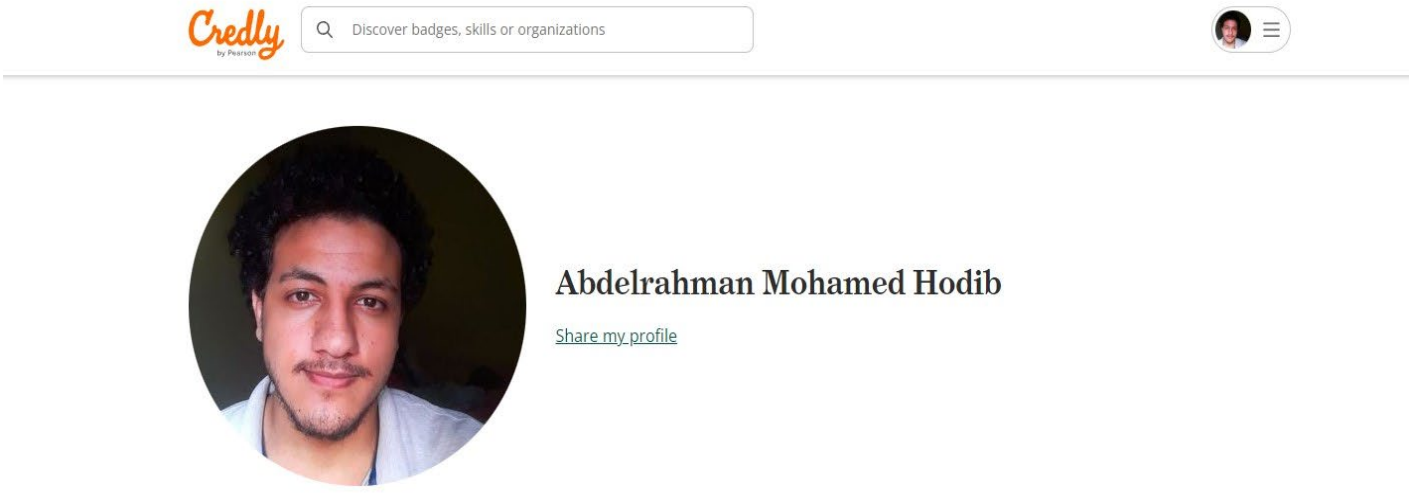




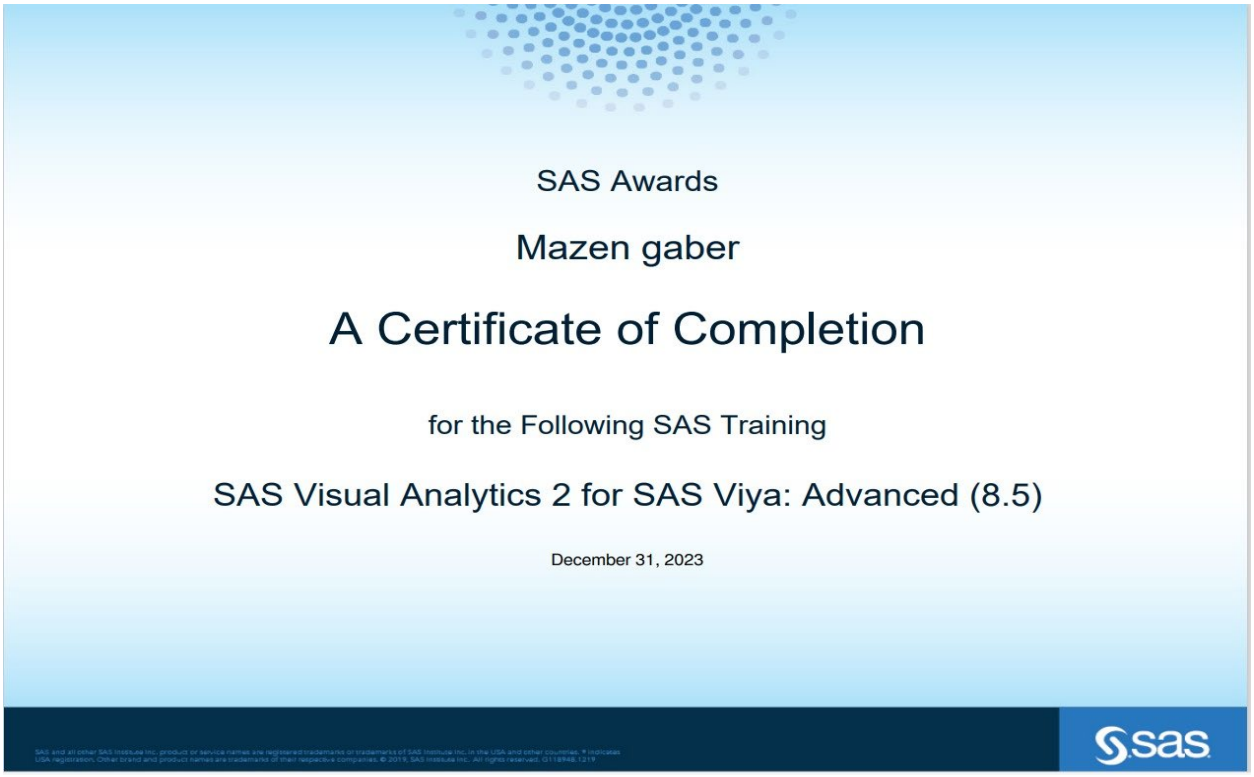
1.2 Abdelrhman Mohamed Abdelhady Hodib 2022513643




SAS Awards
Abdelrhman Mohamed
A Certificate of Completion
for the Following SAS Training
SAS Visual Analytics 2 for SAS Viya: Advanced (8.5)
December 15, 2023
sas


Credly
by Pearson
Discover badges, skills or organizations
Abdelrahman Mohamed Hodib
Share my profile
2 badges
Sort by: Most Popular
SAS Visual Analytics 1 for SAS Viya: Basics
SAS
SAS Visual Analytics 2 for SAS Viya: Advanced
SAS

1.3 Mazen Gaber Mahmoud 20221372110



Mazen gaber

Share my profile

2 badges

Sort by: **Most Popular**

SAS Visual Analytics 1 for SAS Viya: Basics
SAS

SAS Visual Analytics 2 for SAS Viya: Advanced
SAS

1.4 Ahmed Ehab El-Sayed 20221445838

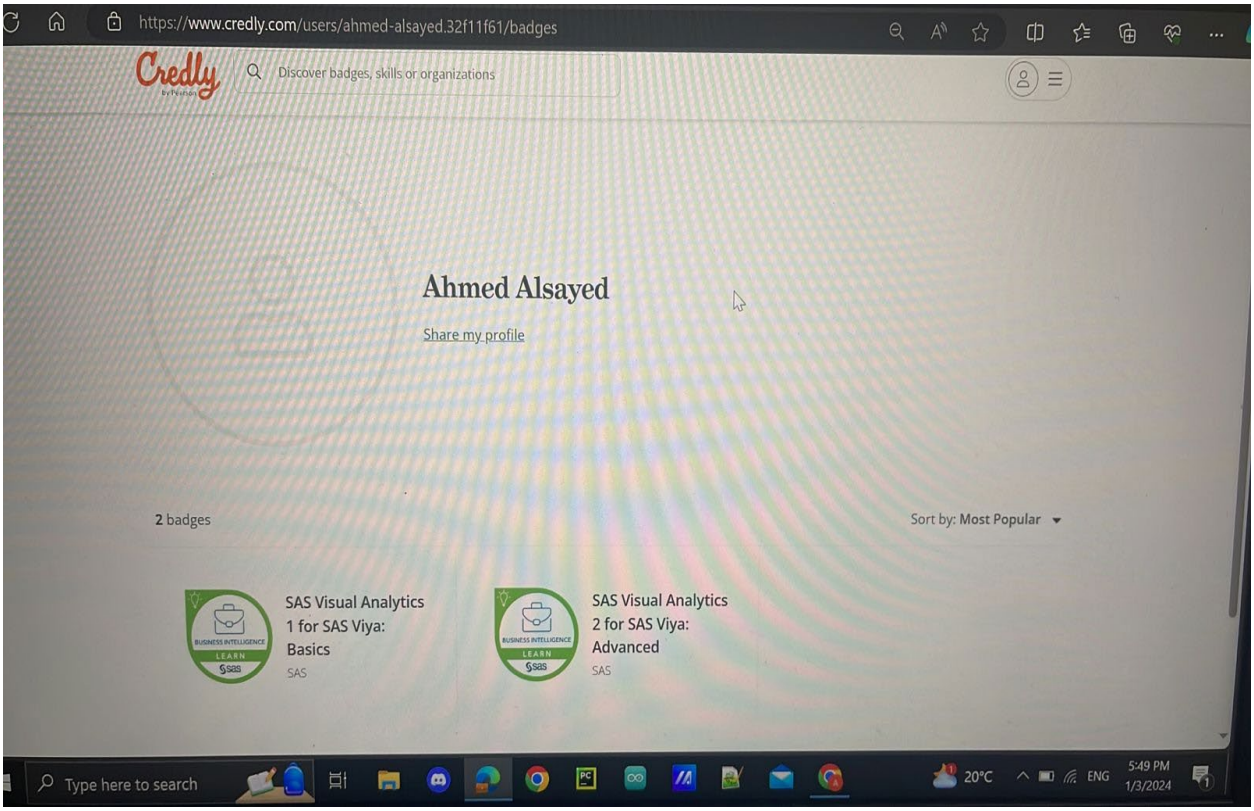
SAS Awards

Ahmed Ehab

A Certificate of Completion

for the Following SAS Training

SAS Visual Analytics 2 for SAS Viya: Advanced (8.5)



2 Introduction

In this Project we will be applying the SAS Visual Analytics Methodology on the AMES_HOUSING dataset on SAS Drive

3 SAS Visual Analytics Methodology

The Visual Analytics methodology is a step-by-step process that we can follow when using Visual Analytics. The methodology is divided into five phases: Access, Investigate, Prepare, Analyze, and Report.

3.1 Access

In the Access phase, we identify the analysis tables that will be used in Visual Analytics and load those tables into CAS.

Accessing the AMES_HOUSING dataset, it consists of 10 Columns and 300 Rows

Here is the breakdown of the variables:

- **Lot_Area**: The area of the lot that the house is on.
- **House_Style**: The style of the house (e.g., 1Story, 2Story, 1.5Fin, SLvl, SFoyer, 1.5Unf, 2.5Unf).
- **Heating_QC**: The heating quality and condition (e.g., Ex for Excellent, Gd for Good, TA for Average/Typical, Fa for Fair).
- **Central_Air**: Whether the house has central air conditioning (Y for Yes, N for No).
- **Gr_Liv_Area**: The above-ground living area in square feet.
- **Fireplaces**: The number of fireplaces in the house.
- **Garage_Area**: The size of the garage in square feet.
- **SalePrice**: The sale price of the house.
- **Age_Sold**: The age of the house at the time it was sold.
- **Season_Sold**: The season when the house was sold (e.g., 1 for Winter, 2 for Spring, 3 for Summer, 4 for Fall).

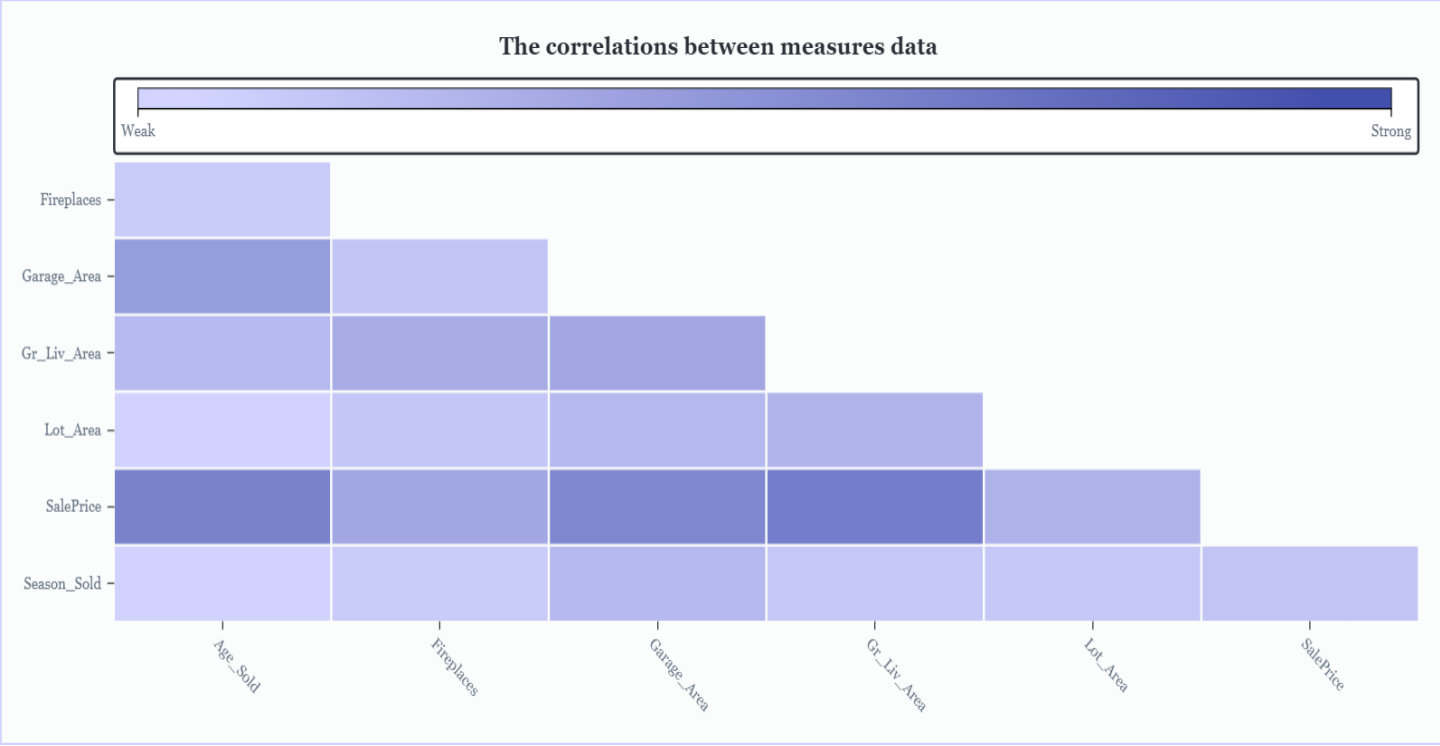
3.2 Investigate

In the Investigate phase, we inspect the tables to determine whether any changes are needed for data items due to data inconsistencies or data quality issues, and we also identify any new data items that need to be created.

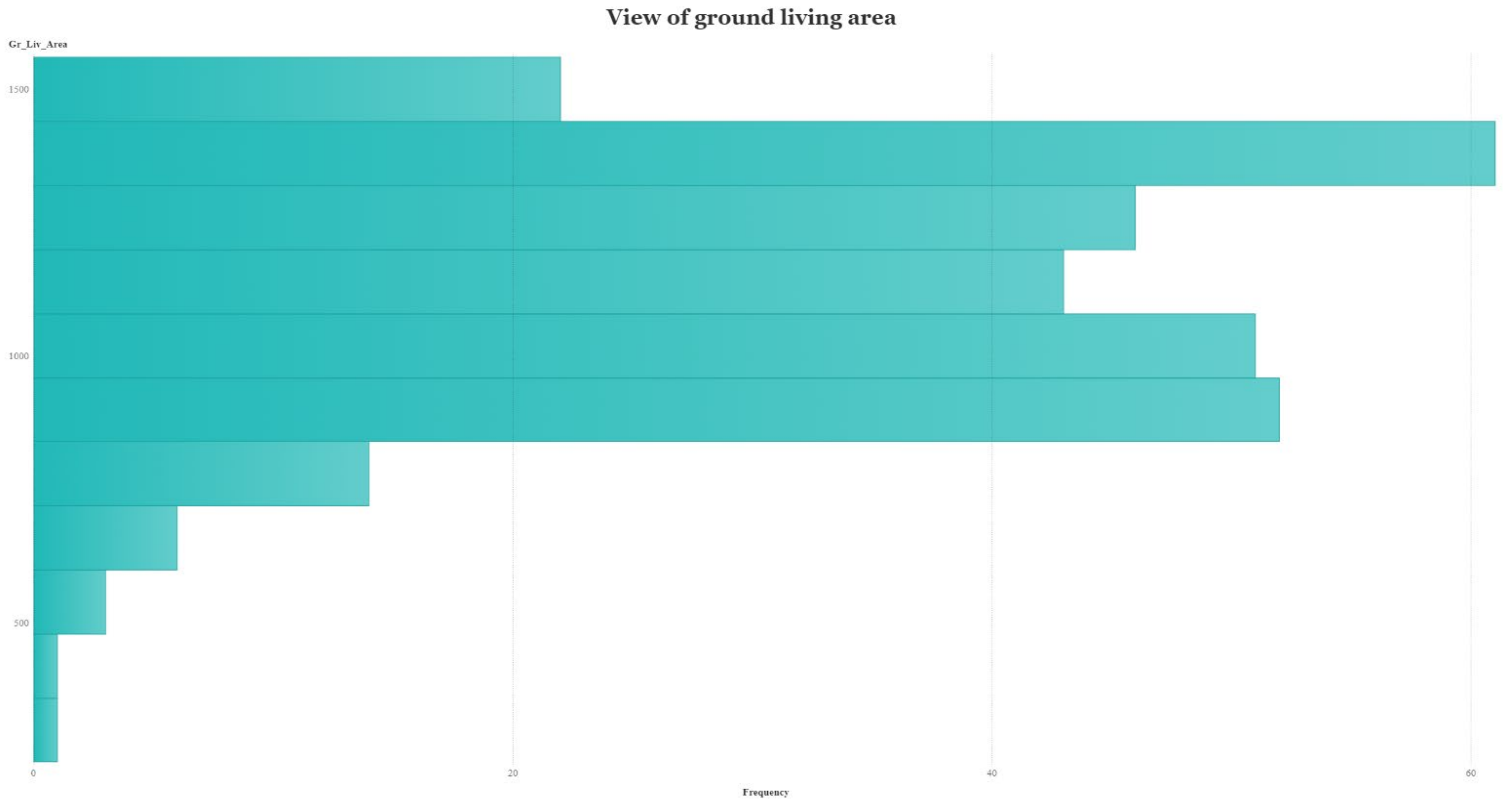
1. We deduced that our data is clean and there are no null values in it.

Column	Unique	Null	Blank	Pattern Count	Mean	Medi
Age_Sold	31.33% (...				45.89	45.
Central_Air	0.67% (2)			1		
Fireplaces	1.00% (3)				0.39	0.
Garage_Area	43.33...				369.45	390.
Gr_Liv_Area	75.67% (2...				1,130.74	1,135.
Heating_QC	1.33% (4)			2		
House_Style	2.33% (7)			4		
Lot_Area	83.33% (250)				8,294.14	8,265.
SalePrice	65.33% ...				137,5...	135,0.
Season_Sold	1.33% (4)				2.55	3.

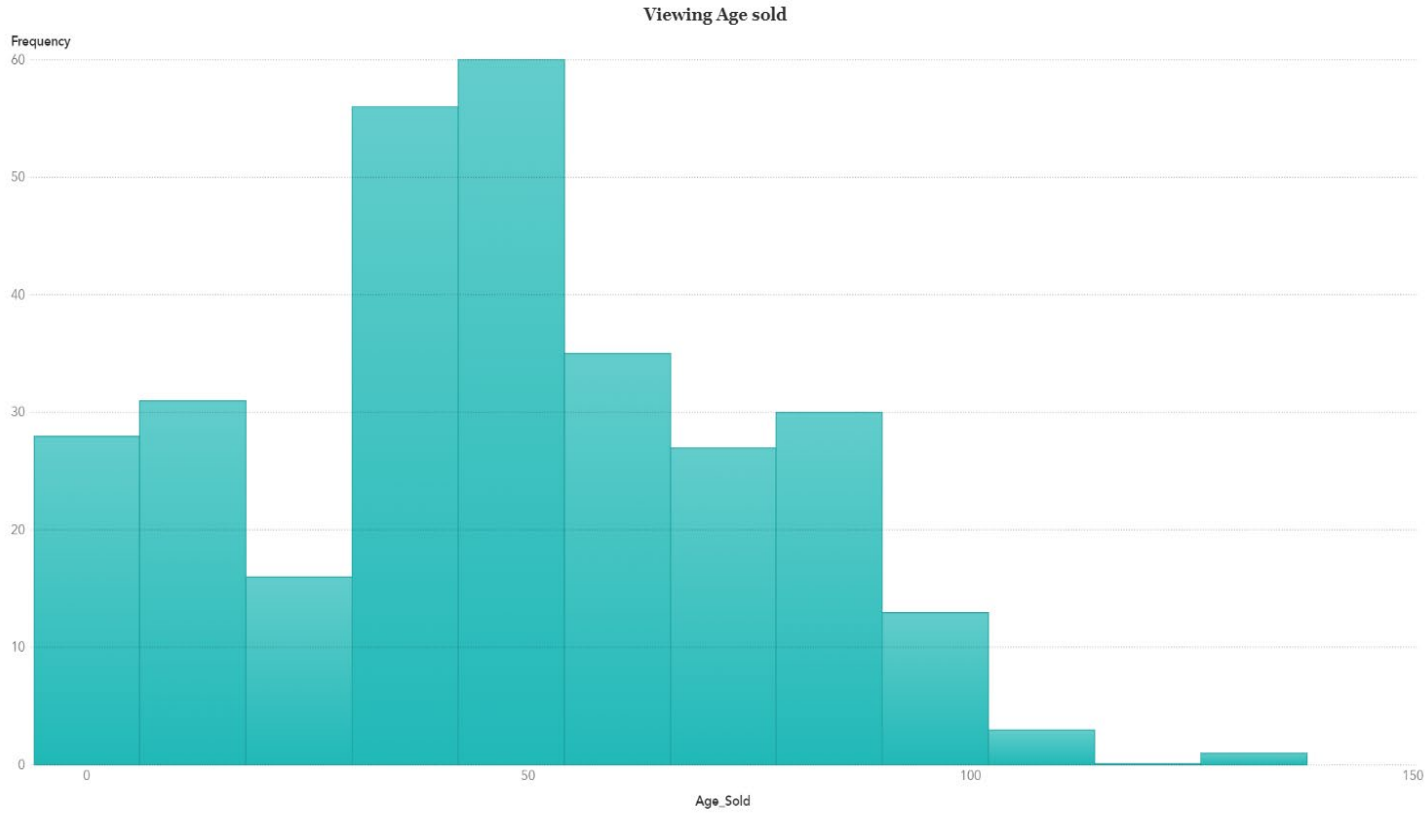
2. We deduced from the correlation matrix between all measures that there is a strong relationship between saleprices and (ground living area, Age sold).



3. We can add new category of ground living area of three classes (less than 500,from 500 to 1000, from 1000 to 1500).



4. We deduced that we can add new data category from the age sold measure that ranges from 0 to 135 into 4 classes (less than 10, from 10 to 30, from 30 to 50 , More than 50).



3.3 Prepare phase

In the Prepare phase, we correct any data quality issues and create any new calculated items needed for analysis.

- We deduced that we can add new data category from the saleprice that has 3 classes (under 100,000, between 100,000 and 200,000, over 200,000).

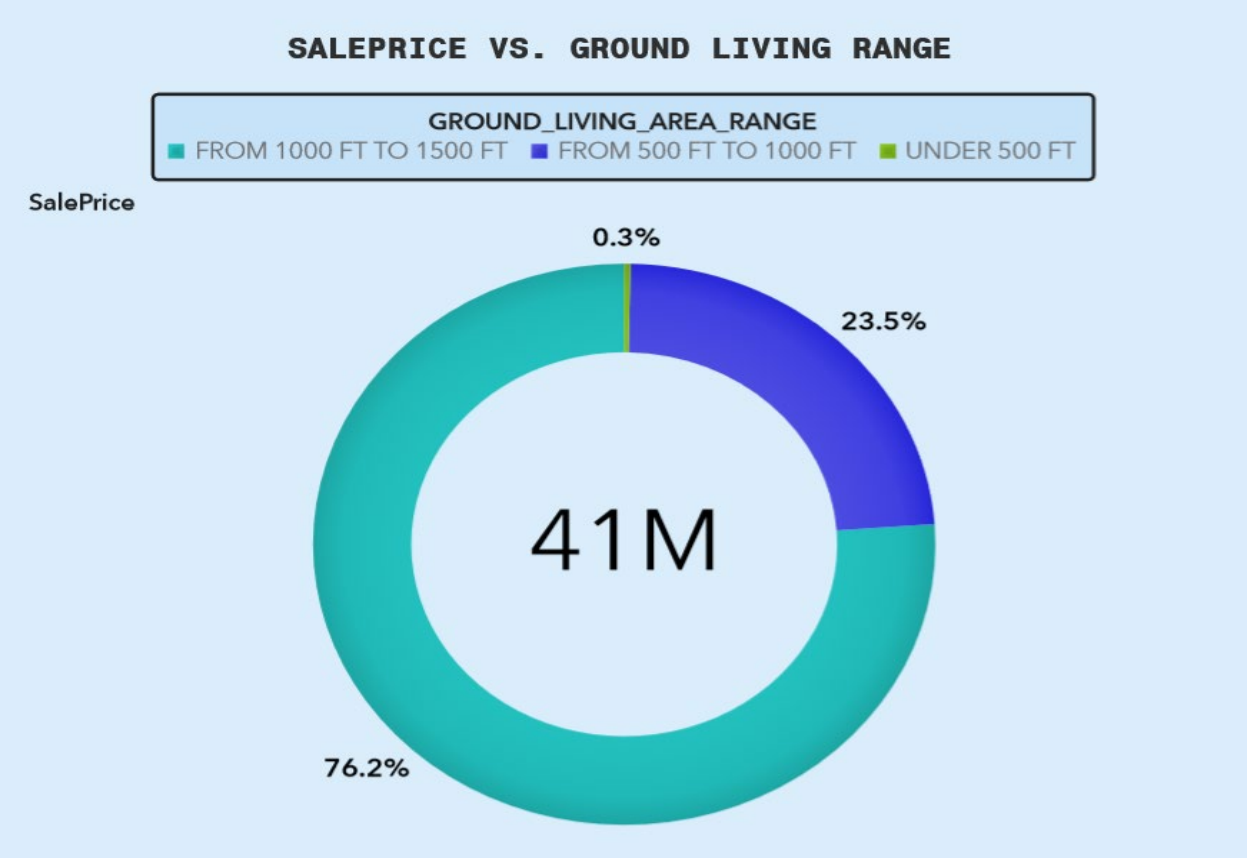
- We deduced that we can change the measure (season_sold) from (1,2,3,4) to (winter,spring,summer,fall).
- We can convert heating quality condition from (ex,fa,…) to Numbers.

3.4 Analyze phase

In the Analyze phase, we explore the data to identify any patterns, relationships, and trends.

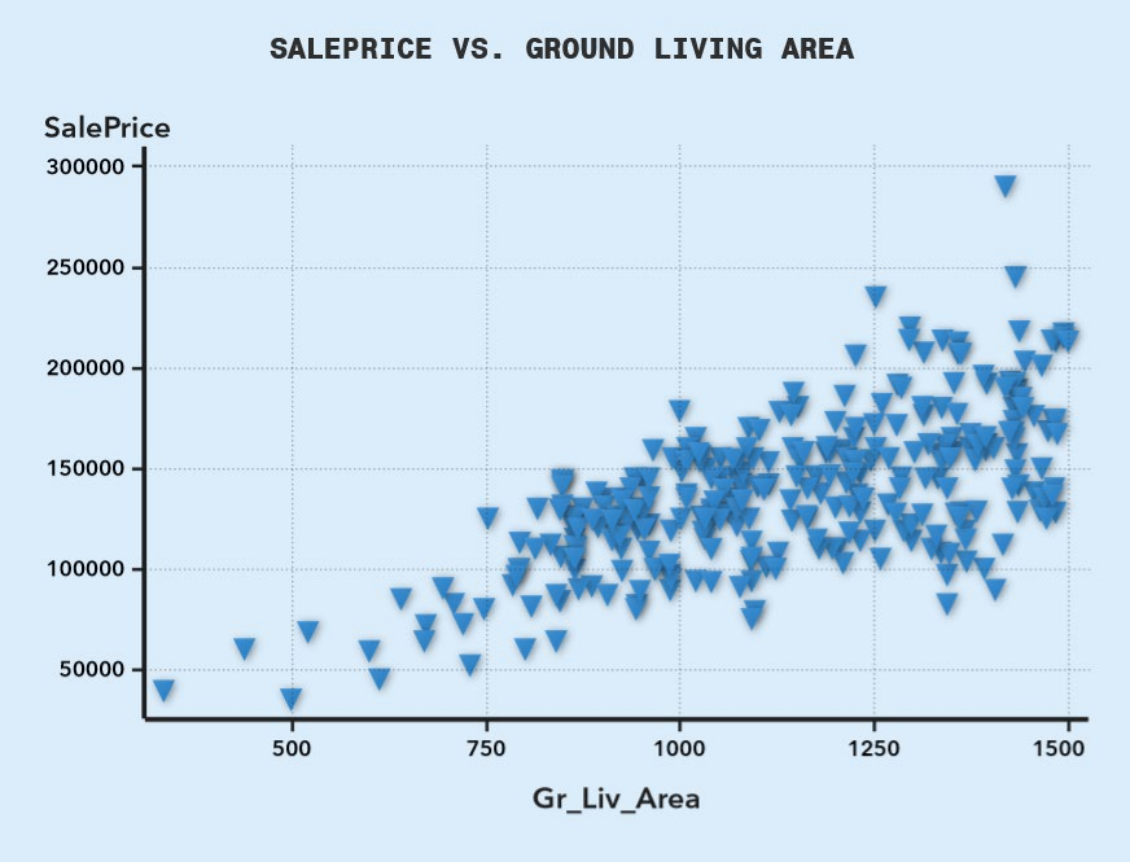
Now, we'll explore our data with charts and graphs to help us uncover hidden relationships.

3.4.1 GROUND LIVING AREA AND SALEPRICE



The pie chart shows the distribution of sale prices for the ranges of the ground living area.

- Houses with a ground living area range of 1000 sq ft to 1500 sq ft make up the largest portion , at 76.2%.
- The next largest group is houses with a ground living area range of 500 sq ft to 1000 sq ft, at 23.5%.
- The smallest group is houses with a ground living area under 500 sq ft, at 0.3%.

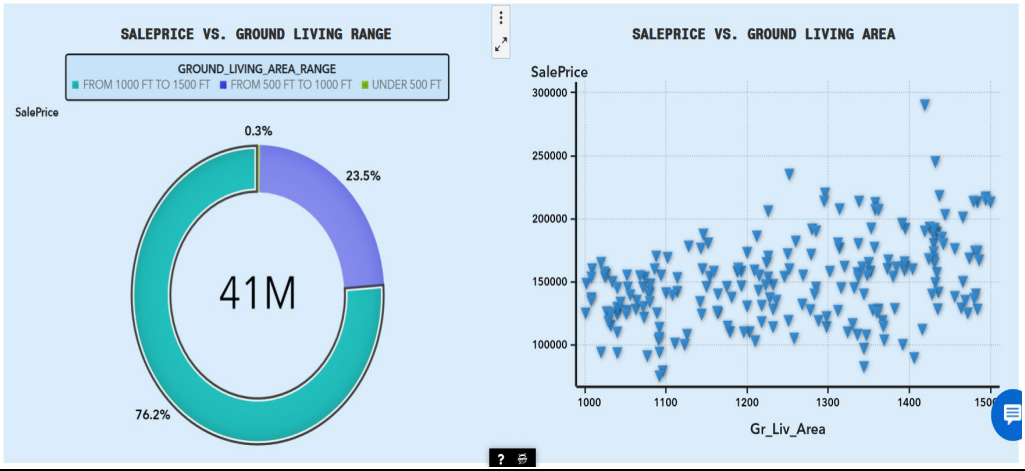
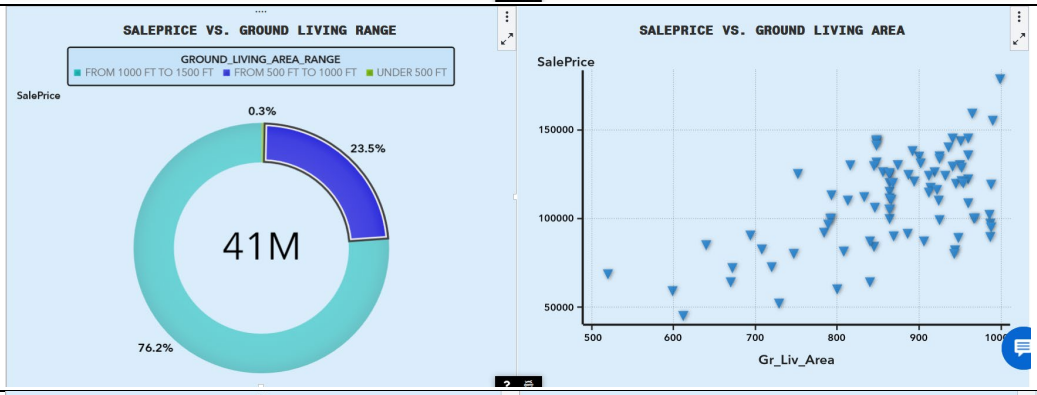
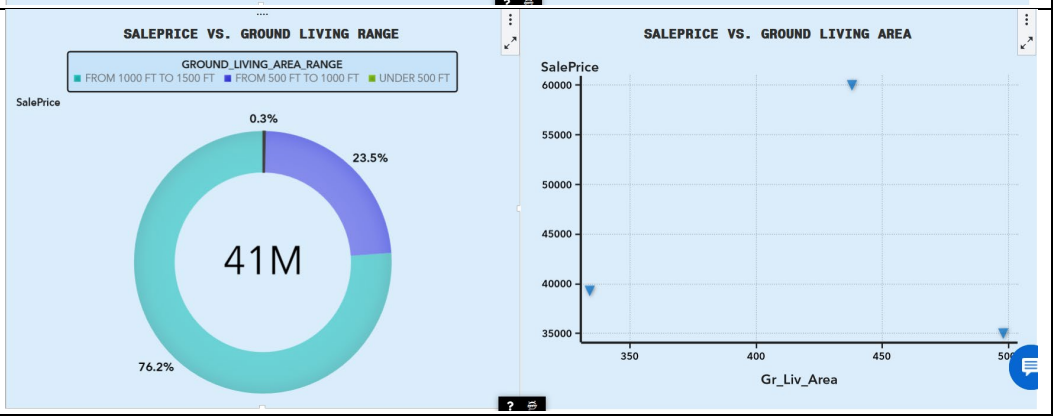


The scatter plot shows the difference in sale price for ground living area of houses.

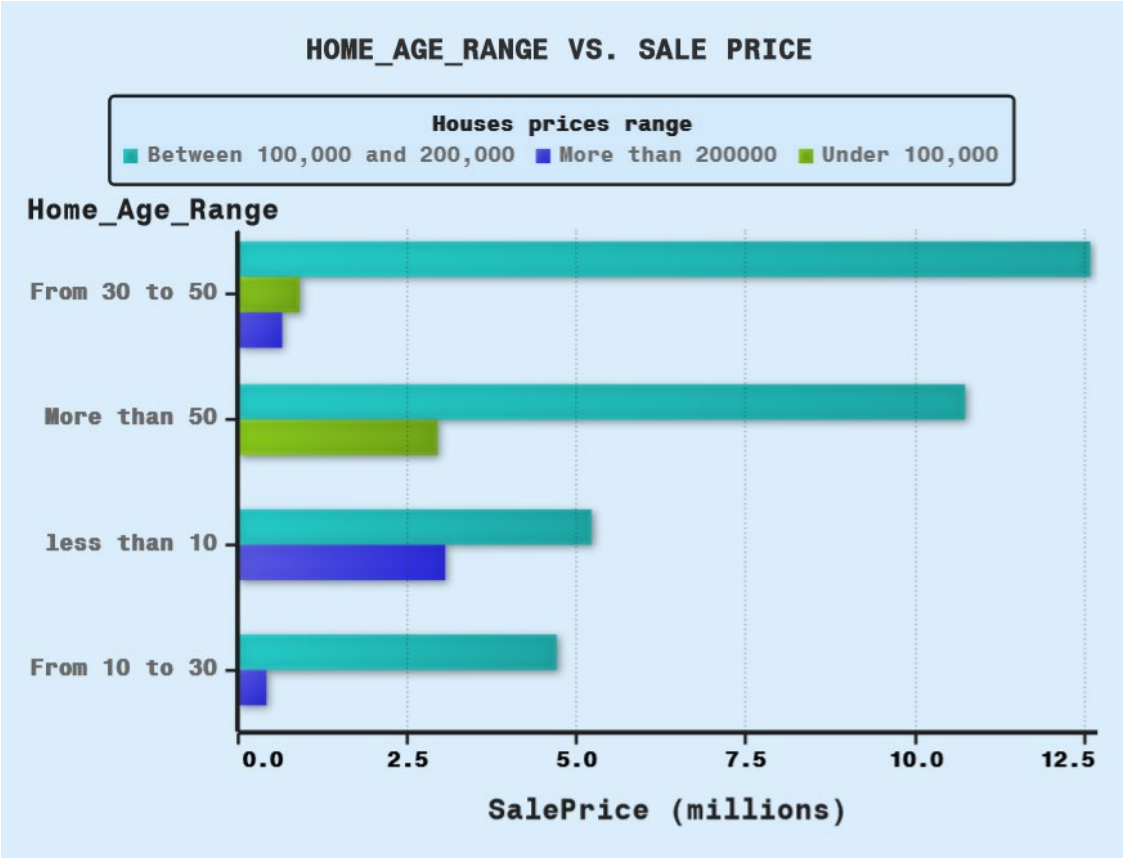
- The general trend of the graph is that houses with larger ground living areas tend to have higher sale prices. This is because more

space is generally seen as a desirable feature in a house, and buyers are willing to pay more for it. However, there is also a lot of variation in the data, and there are many houses that fall below or above the general trend.

Here is the view of sale prices regarding each range of the ground living area

FROM 1000 FT TO 1500 FT	
FROM 500 FT TO 1000 FT	
UNDER 500 FT	

3.4.2 HOME AGE AND SALEPRICE



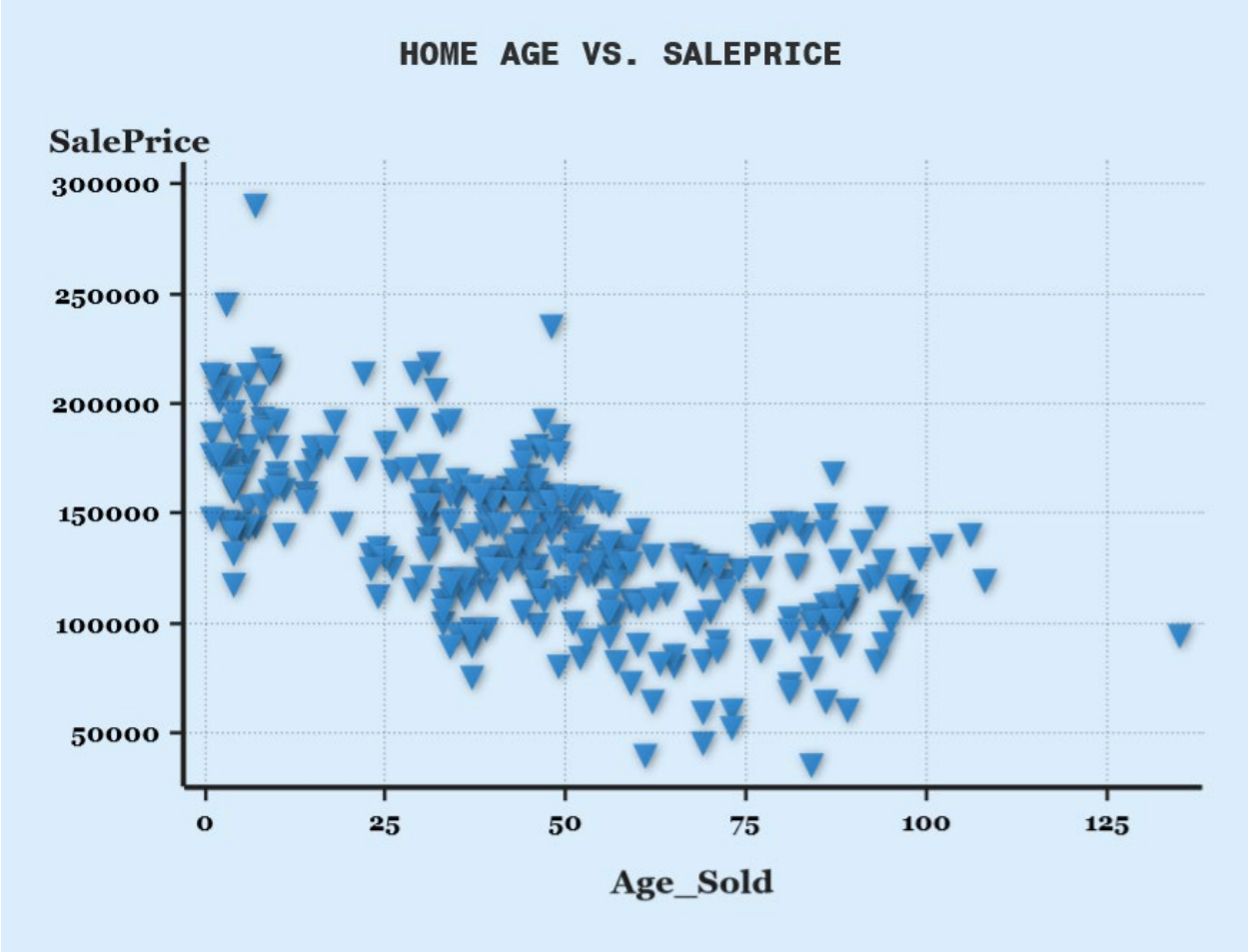
The bar plot shows the sale price of houses by age range.

The graph shows that houses between 30 and 50 years old have the highest median of all saleprice, at around $7.5 million. Houses from 10 to 30 years old have the lowest median of all sale price, at around $2.5 million. Houses more than 50 years old have a median of all sale price of around $5 million.

There are a few possible explanations for this pattern.

- One possibility is that newer houses are more expensive because they are built with newer materials and technologies.

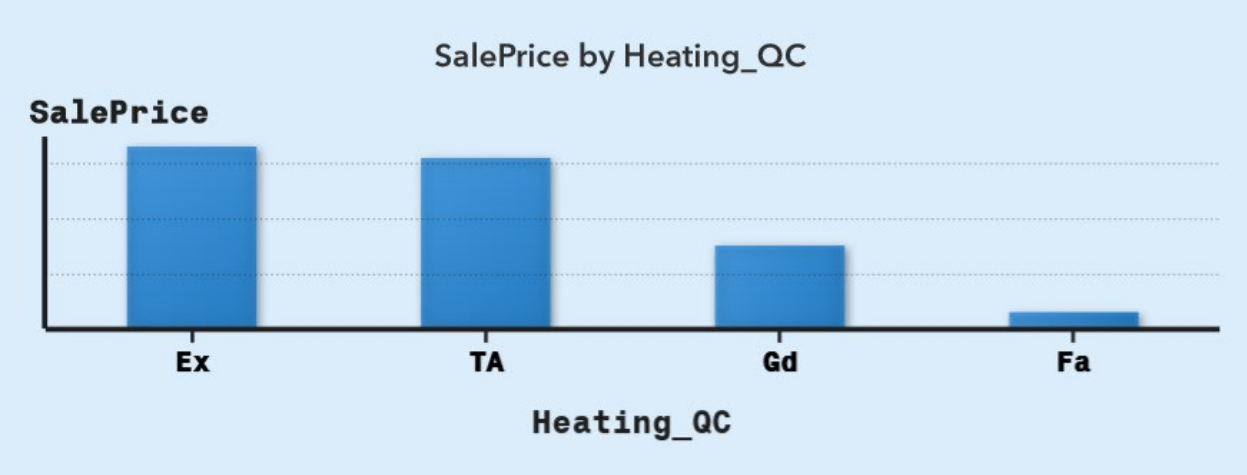
- Another possibility is that older houses are more sold because they are located in desirable neighborhoods. It is also possible that there is a combination of these factors at play.



The scatter plot shows the distribution of sale price of homes by the age of the homes were sold.

- The graph shows that, in general, newer homes tend to have higher salseprice for more than older homes.

3.4.3 HEATING QC AND CENTRAL AIR ON SALE PRICE



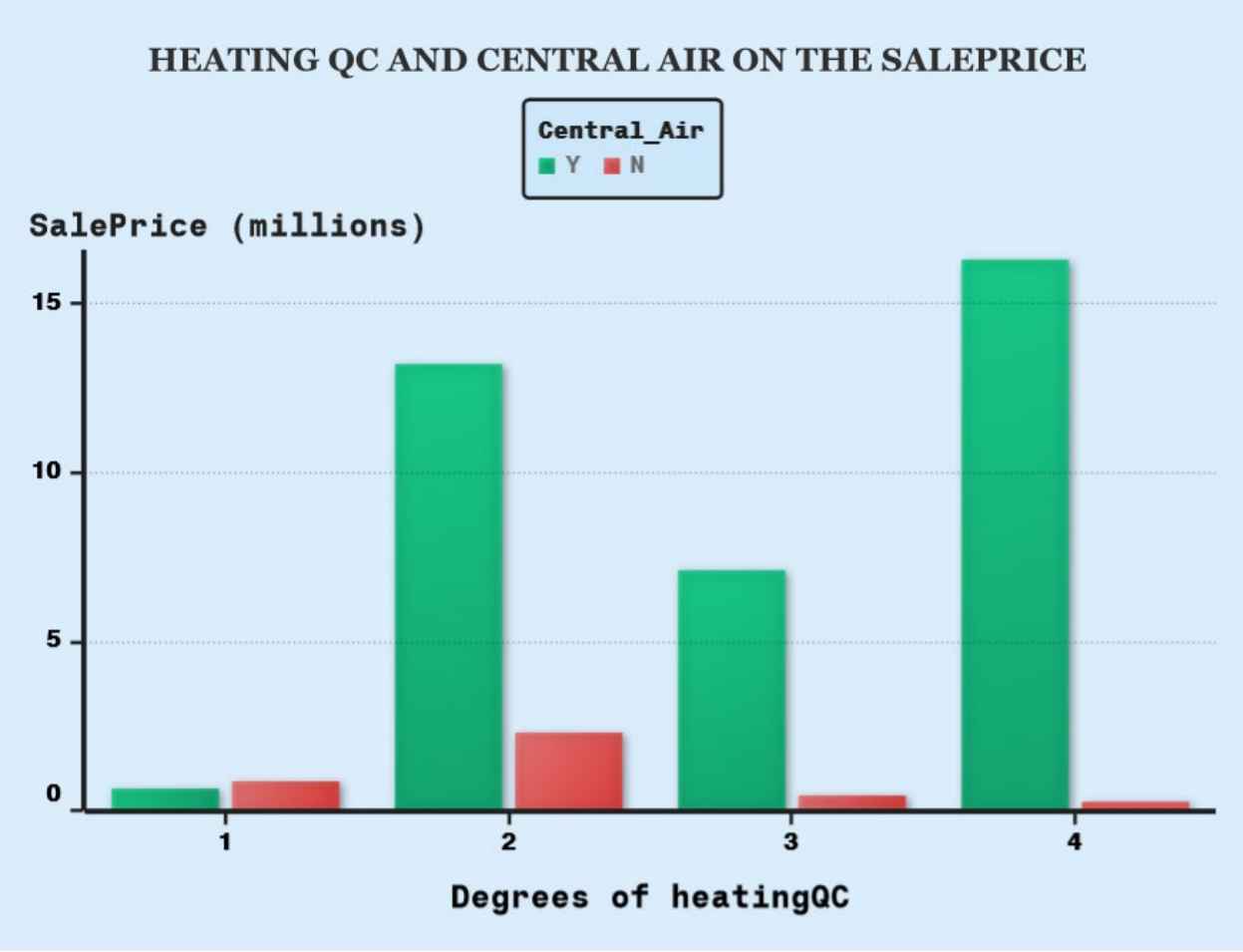
The bar plot shows sales price of homes, by heating quality condition (heating_qc)

- Homes with excellent heating quality condition have the highest sum of sales price(16.6 million).
- Homes with typical average heating quality condition have slightly lower sum of sales price than excellent ones(15.5 million).
- Homes with good heating quality condition have a lower sum of sales price at around 7.6 million.
- Homes with fair heating quality condition have the lowest sum of sales price, at around 1.5 million.

The bar plot shows sales price of homes, by Central Air conditioning

It is clear that houses with Central Air conditioning have higher sum of sale prices than those without

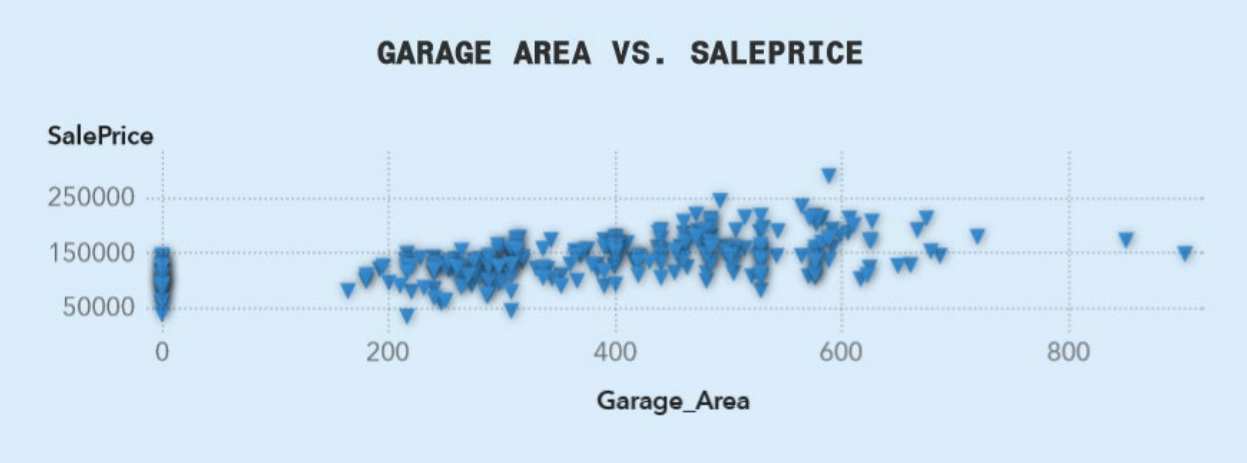
- Central air conditioning provides consistent cooling throughout the home, enhancing comfort, especially in warmer climates or during hot seasons. Buyers often prioritize this feature, leading to increased demand and higher prices for homes equipped with central air conditioning.



The graph shows the relationship between the sale price of homes (in millions) and the number of degrees of heating and Central Air conditioning the home has.

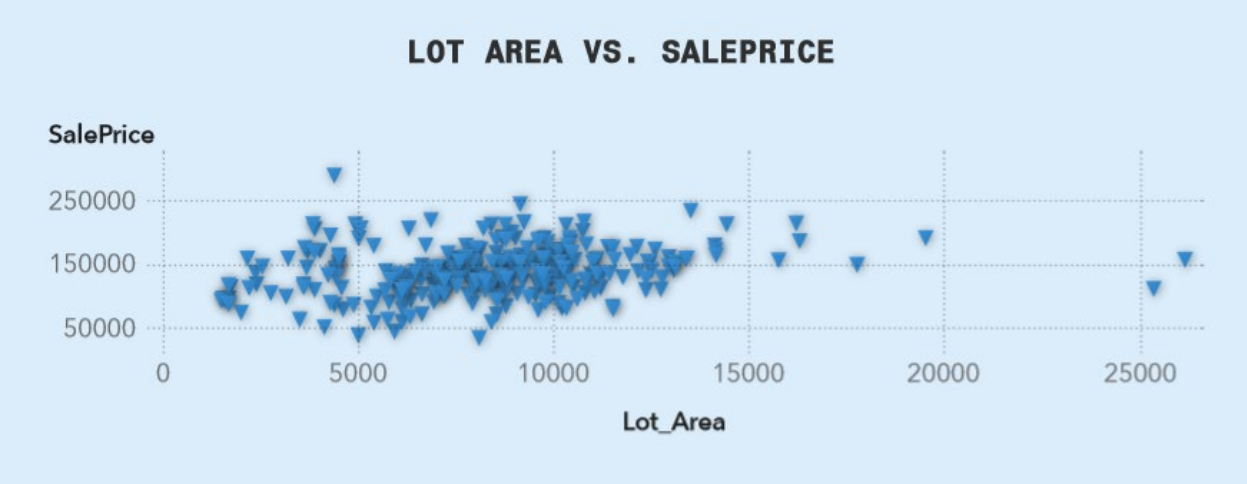
As we can see from the graph, there is a positive correlation between the sale price of a home and the number of degrees of heating and Central Air conditioning the home has. This means that homes with more heating and cooling tend to sell for more money as they are more comfortable to live in, which can make them more appealing to buyers.

3.4.4 GARAGE AREA AND LOT AREA ON HOUSE PRICE



The scatter plot shows the relationship between the size of a garage and the sale price of the house it's attached to.

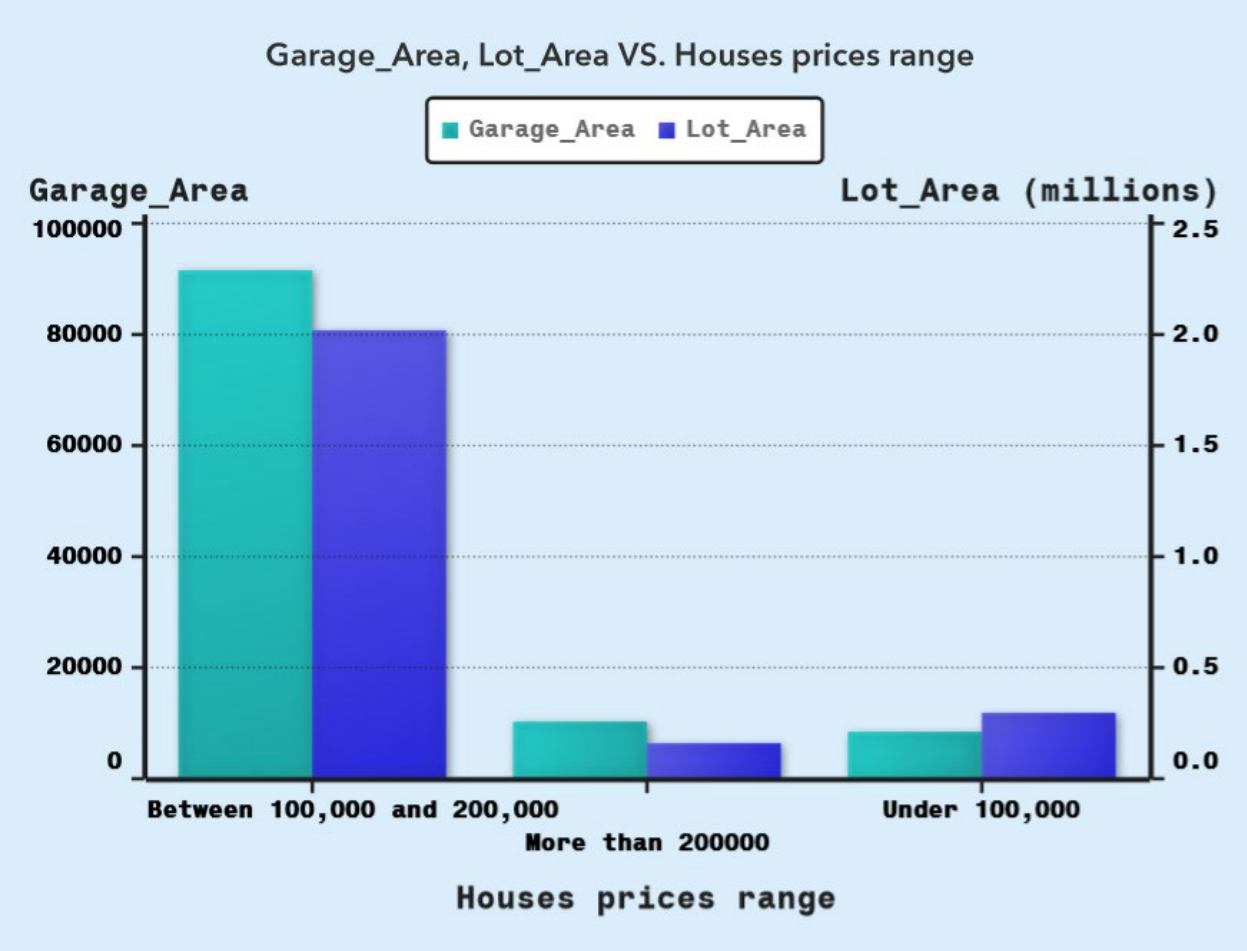
It suggests that there is a positive correlation between the size of a garage and the sale price of a house. This means that houses with larger garages tend to sell for more money than houses with smaller garages.



The scatter plot shows relationship between lot area and sale price for houses.

The graph shows that there is a positive correlation between lot area and sale price. This means that houses with larger lots tend to sell for more money than houses with smaller lots. There are a few possible explanations for this.

- A larger lot gives homeowners more space to enjoy, such as for a larger yard, a pool, or a garden. This can make the house more desirable to buyers and willing to pay more for it.
- A larger lot may also give homeowners the potential to develop the property further, such as subdividing the lot. This can also make the house more valuable.
- In some cases, larger lots may be located in more desirable areas, which can also contribute to a higher sale price.

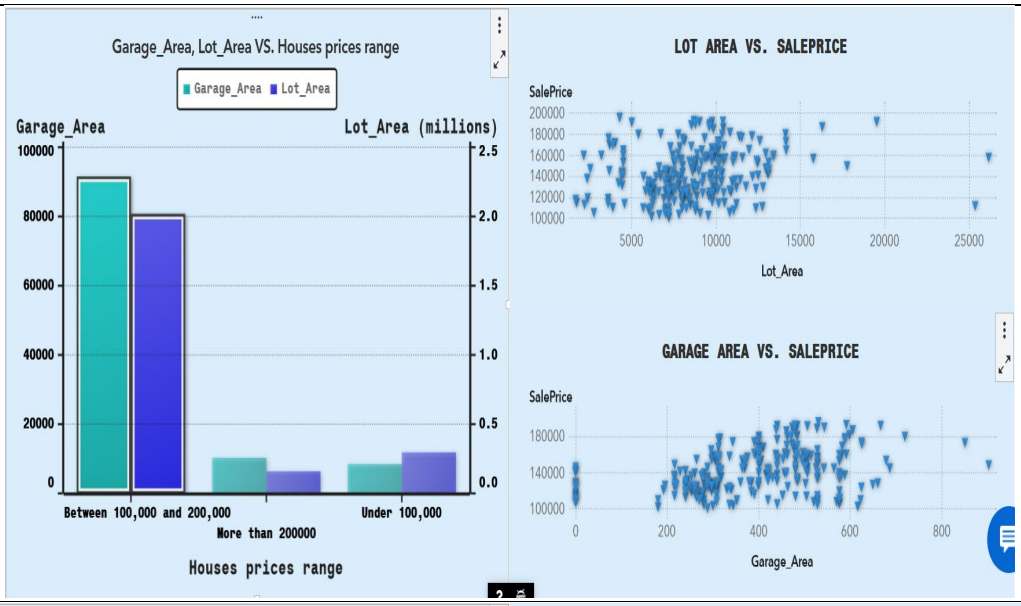
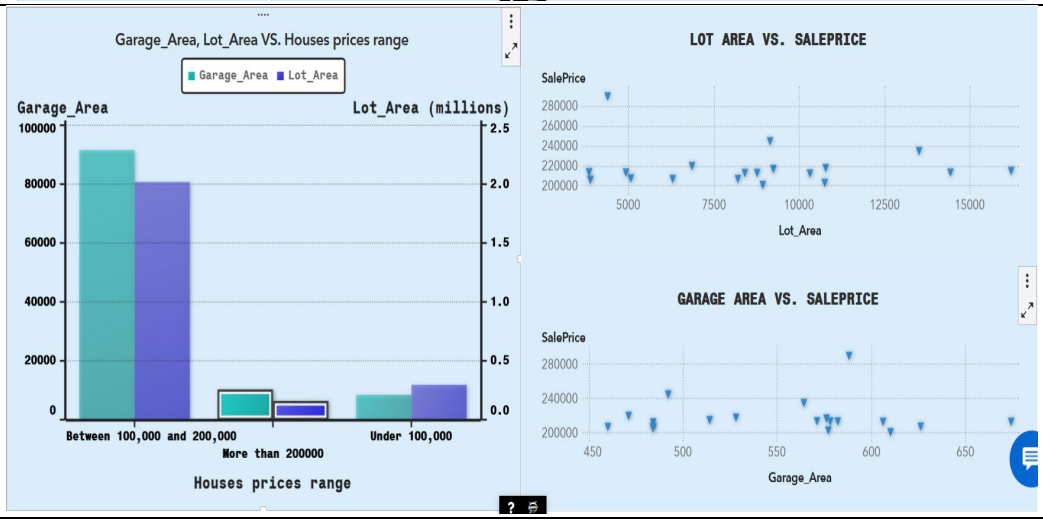


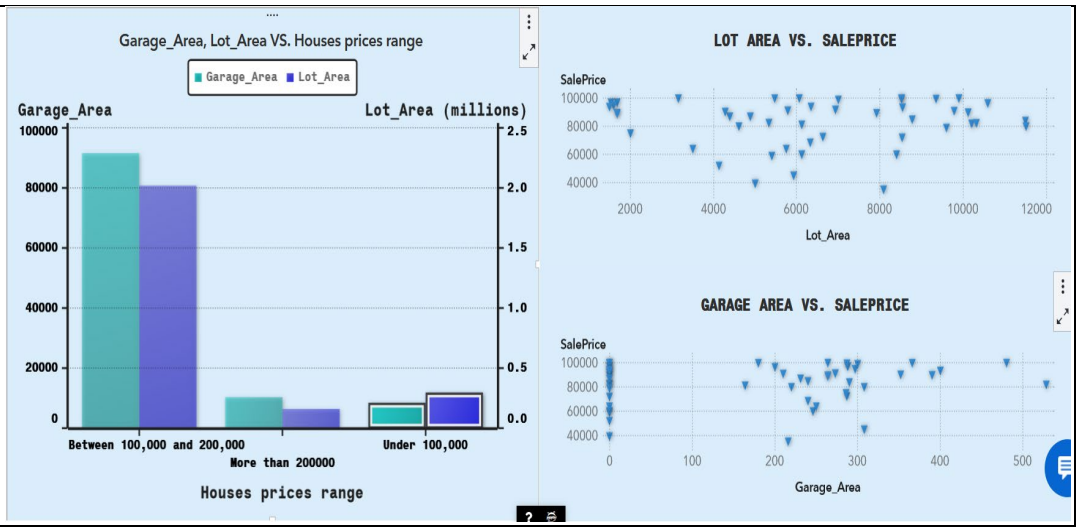
From that plot we could deduce that houses their prices range between 100,000 and 200,000, have a highest summation of garage area and lot_area.

This ensures the positive correlation between both the size of a garage and lot area with the sale price of a house

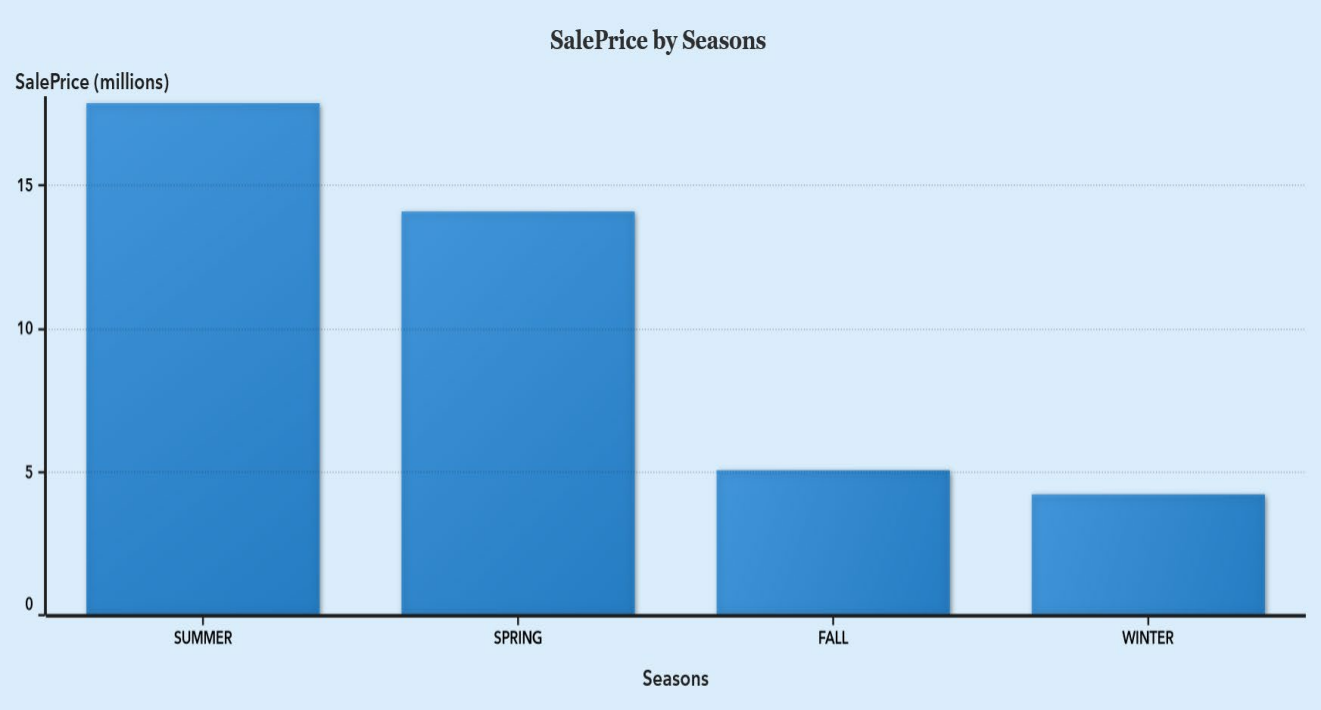
Also for houses more than 200,000 , the Garage Area seems to be more important than lot area.

Here is the distribution of both the size of a garage and lot area on sale price ranges

Between 100,000 and 200,000	
More than 200,000	

Under 100,000	

3.4.5 SEASONS AND SALEPRICE



Regarding the bar plot between sale price for houses by season :

- The summation of sale price for houses in the summer is the highest.
- The summation of sale price for houses in the winter is the lowest.
- The summation of sale price for houses in the spring is higher than the sale price for houses in the fall

Here are some possible explanations for that:

- Pleasant weather: Spring and summer are more attractive times to buy a house with pleasant weather for outdoor activities like house viewings and landscaping.
- Schooling: Families with children might prefer to move during the summer to minimize disruption to their schooling.

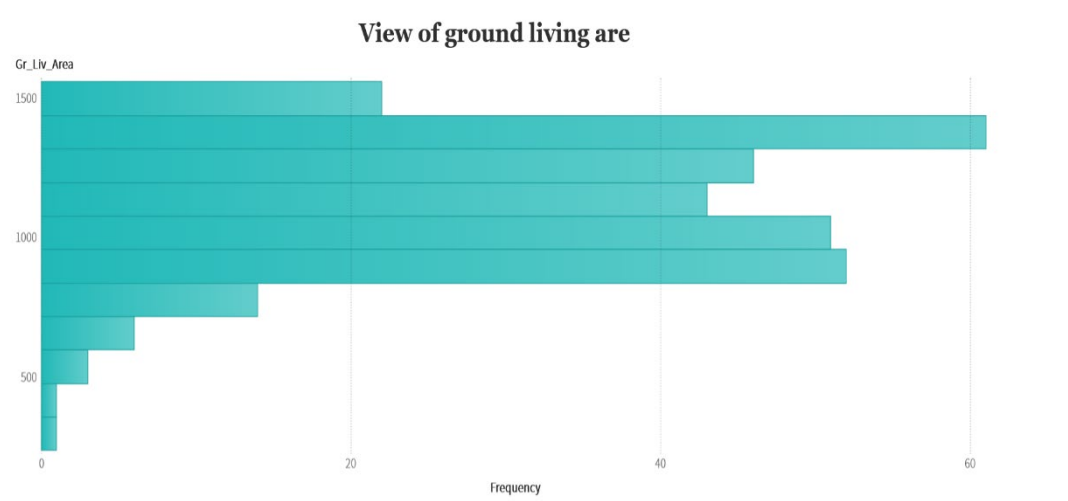
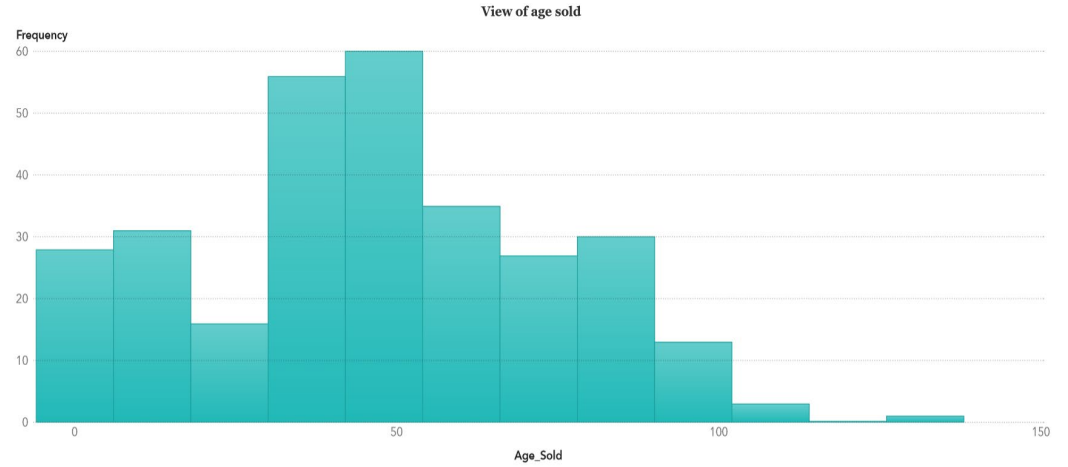
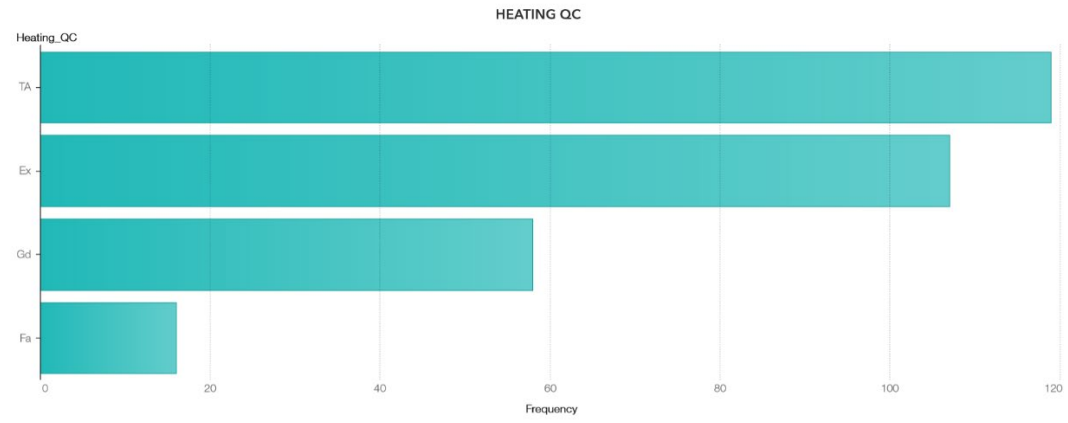
3.5 Report

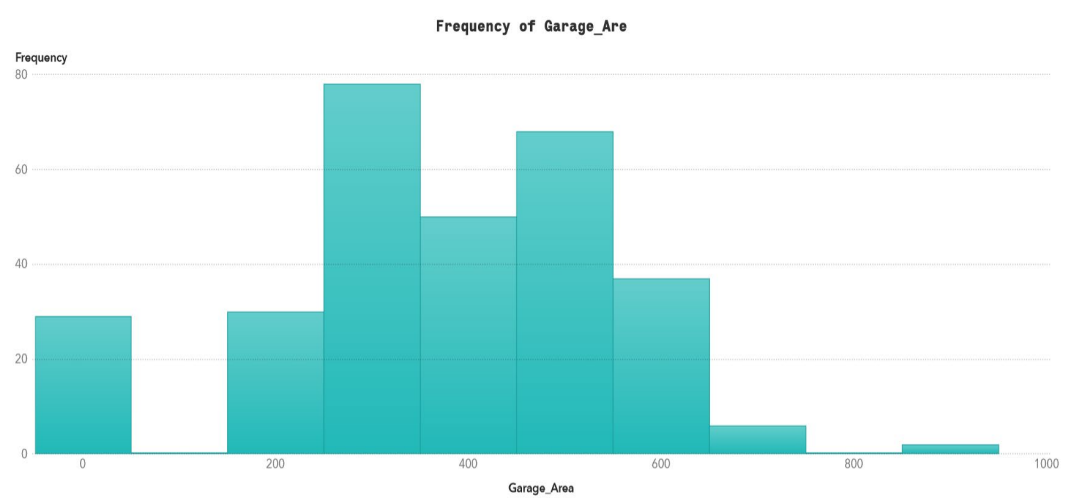
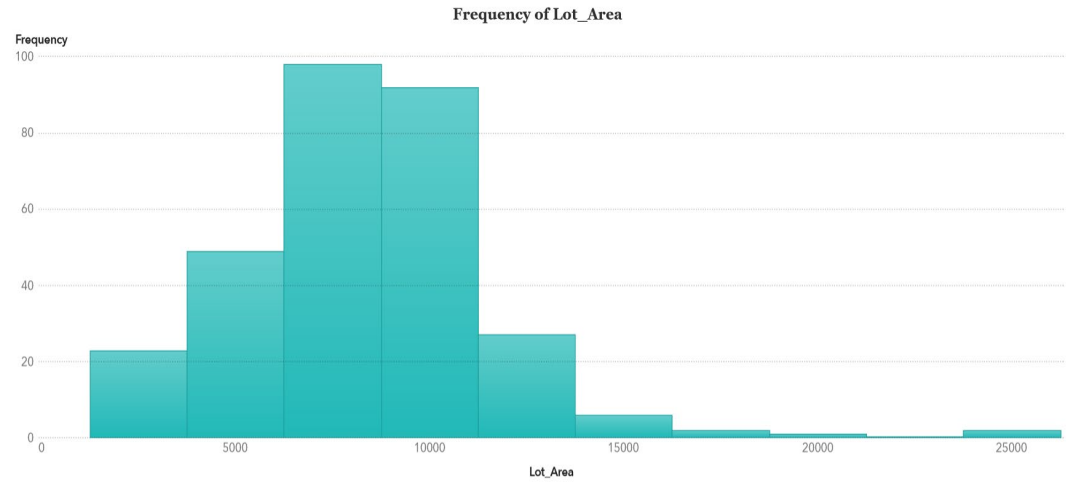
In the Report phase, we develop interactive reports that can be shared via the web or a mobile device.

- Hidden pages

Hidden pages are a feature of SAS Visual Analytics that allow us to include a data item in the query results when we want its values, but do not want it to be displayed

Regarding our report we have made 5 hidden pages to enable viewers to see additional details.

Ground Living Area	
Age Sold	
Heating QC	

Garage Area	
Lot Area	

Report video : https://drive.google.com/drive/folders/1L-P459a1fvGl8CUHlkQfEyA8sbF44MZX?usp=sharing