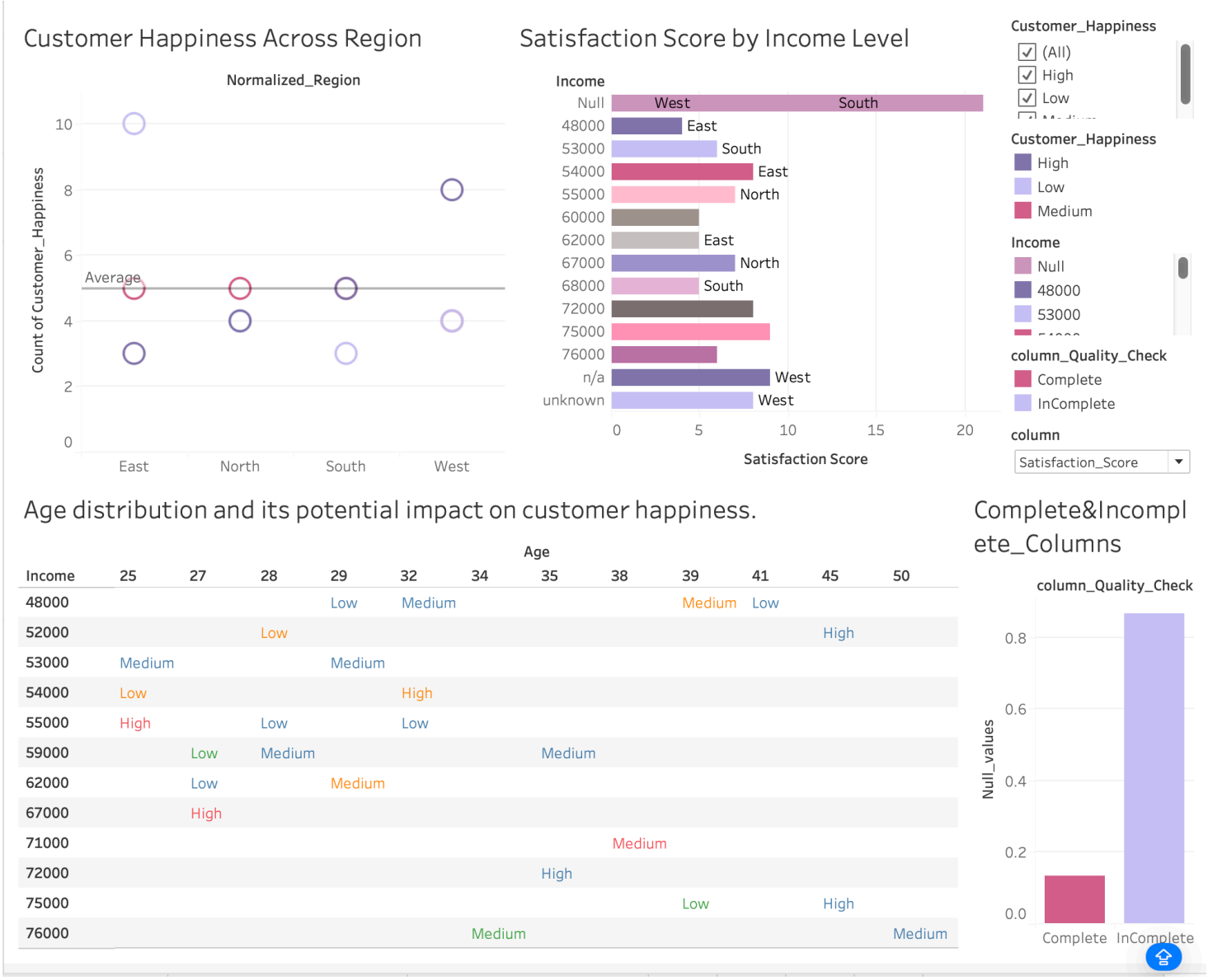
Summary Report



Income	25	27	28	29	32	34	35	38	39	41	45	50
48000				Low	Medium				Medium	Low		
52000			Low								High	
53000	Medium			Medium								
54000	Low				High							
55000	High		Low		Low							
59000		Low	Medium				Medium					
62000		Low		Medium								
67000		High										
71000								Medium				
72000							High					
75000									Low		High	
76000						Medium						Medium

- It is observed that a customer earning $75,000 has a high satisfaction score of 9, but their region is not specified.
- Another customer with an income of $76,000 has a satisfaction score lower than 9.
- A customer earning $54,000 has a satisfaction score of 8, showing that satisfaction scores do not directly correlate with income levels.
- Two customers, one aged 25 and the other 32, both with the same income level, are in different happiness segments, Low and High respectively.
- A 25-year-old with an income of $55,000 is highly happy.

- This indicates that several factors are missing in the dataset to identify the customer's happiness and satisfaction scores.

Data Quality Check:

Null Value Summary:

- **Region**: 60 missing values
- **Age**: 3 missing values, 55 non numeric values- error, unknown
- **Income**: 43 missing values, 33 non numeric values- error, unknown
- **Satisfaction Score**: 104 missing values

Missing Region Handling:

There are 60 entries with missing values in the Region column. These missing entries could be handled by:

Assigning a default category like 'Not Present' or 'unknown' to maintain record completeness for certain analyses.

Data Consistency Checks:

- **Region Consistency**: Several rows have incorrect or missing region values. Imposing a consistent format using expected labels ('North', 'South', 'East', 'West') or cleaning the data to fit these categories is recommended.

- **Customer Feedback Consistency**: Feedback entries often do not start with a capital letter or end with proper punctuation, suggesting a need for standardizing text entries to improve textual analysis or machine learning preprocessing.

- To ensure the consistency of the 'Region' and 'Customer Feedback' data fields, we can use the string function named 'Proper'. This function will capitalize the first letter of each word.

Data Cleaning:

- Convert Age and Income fields to numeric, handle non-numeric errors by replacing them with NaN or imputing sensible values.
- Standardize Customer Feedback to begin with a capital letter and end with a punctuation mark.
- In order to correct the fields with missing values, I built calculated fields. Our dataset has missing values in variables such as age, region, income, and acceptable score.

Additional Data Elements:

- **Customer Demographics**: Adding additional demographic information, such as gender and occupation, may help uncover patterns in customer satisfaction.
- **Purchase History**: Information about previous purchases, transaction volume, and average expenditure may be useful in determining customer satisfaction levels.

- **Region-Specific Details**: The study could be improved, particularly in focused marketing efforts, by using more detailed geographical data, such as urban vs. rural.