

Project Background

Car resale is a lucrative business, and many cars in the United States undergo the process of change in ownership. This can be due to various factors, such as the aging technology in the car, interest in obtaining newer car by the present owner, having an extensive ridership history and exhausting its prime period, purchasing cars to use them for or sell them after being used for commercial purposes such as taxis, transport vehicles, or for vintage collectibles as a memory of their historic value.

Problem Statement

Executive Summary

Dataset

Problem Statement

The features and condition of a car are the most important indicators of a car's resale and purchase values, which determine where the price can be increased or where it could depreciate.

Project Background

Car resale is a lucrative business, and many cars in the United States undergo the process of change in ownership. This can be due to various factors, such as the aging technology in the car, interest in obtaining newer car by the present owner, having an extensive ridership history and exhausting its prime period, purchasing cars to use them for or sell them after being used for commercial purposes such as taxis, transport vehicles, or for vintage collectibles as a memory of their historic value.

Problem Statement

Executive Summary

Dataset



Executive Summary

This project tries to provide a visualization model for the data, highlighting key factors such as company, accidents or damages occurred, color , enginetype, drivetrain, and various other factors. This can facilitate applications like purchase feasibility study on the car using the data, and various other functions such as price negotiability, car scrap cost and the repair analysis of the damages on the car.



Project Background

Car resale is a lucrative business, and many cars in the United States undergo the process of change in ownership. This can be due to various factors, such as the aging technology in the car, interest in obtaining newer car by the present owner, having an extensive ridership history and exhausting its prime period, purchasing cars to use them for or sell them after being used for commercial purposes such as taxis, transport vehicles, or for vintage collectibles as a memory of their historic value.

Problem Statement

Executive Summary

Dataset

DATASET

- Data consists of US Used Car details
- Collected from kaggle
- size: 9GB



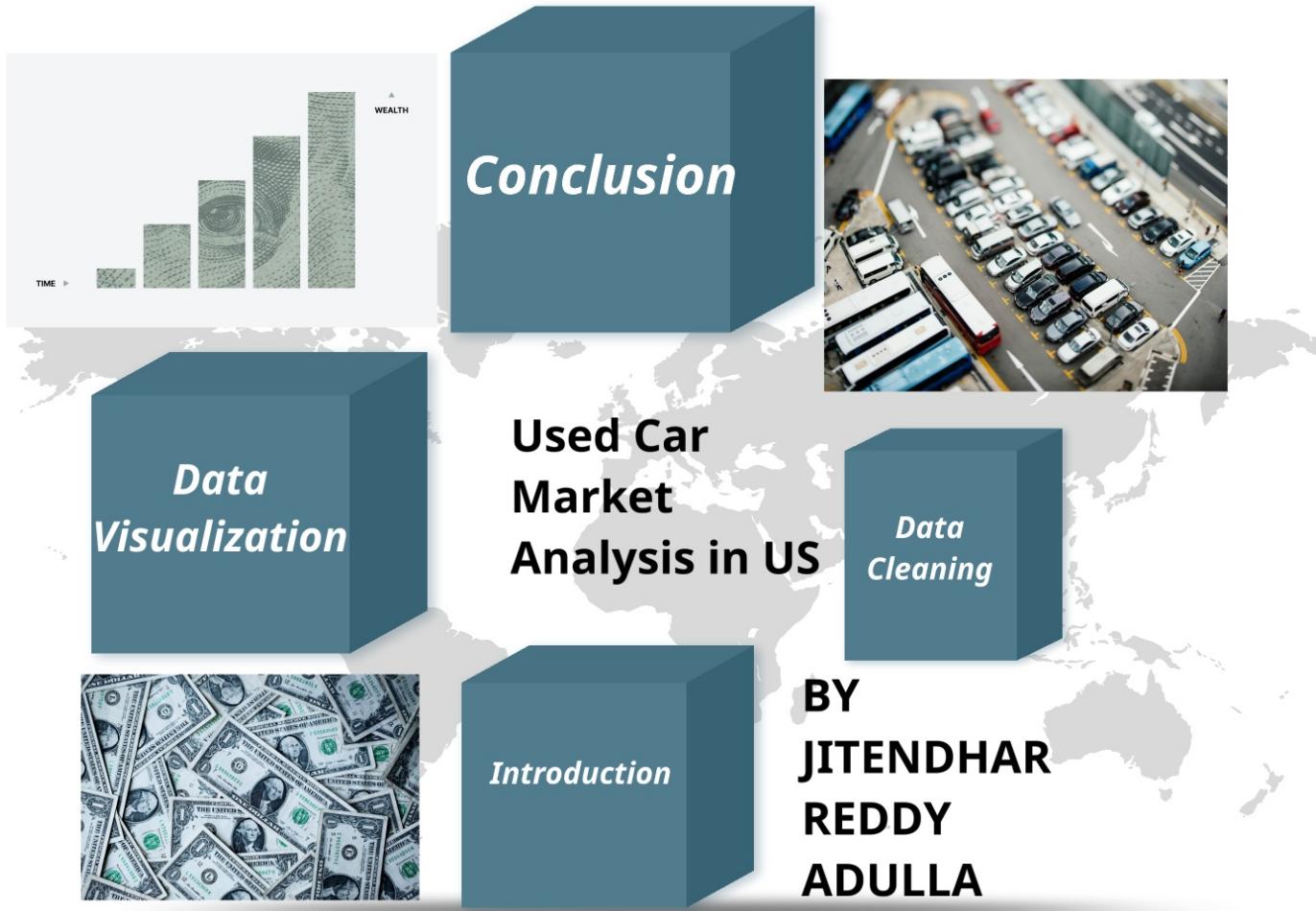
Project Background

Car resale is a lucrative business, and many cars in the United States undergo the process of change in ownership. This can be due to various factors, such as the aging technology in the car, interest in obtaining newer car by the present owner, having an extensive ridership history and exhausting its prime period, purchasing cars to use them for or sell them after being used for commercial purposes such as taxis, transport vehicles, or for vintage collectibles as a memory of their historic value.

Problem Statement

Executive Summary

Dataset



*Replacing
Values*

Data Cleaning

- Null Values
- Replacing values

*Null
Values*

NULL VALUES

- Drop columns with more null values
- Replace Null Values with Mean Value

```
# dropping the columns which has more null values
cars_df.drop(['cabin','bed','bed_height','bed_length','combine_fuel_economy','vehicle_damage_category'], inplace=True)

# cars_df.columns
Index(['vin', 'back_legroom', 'body_type', 'city', 'city_fuel_economy',
       'days_in_stock', 'diameter_rim', 'description', 'engine_cylinders',
       'engine_displacement', 'engine_type', 'exterior_color', 'fleet',
       'frame_damaged', 'franchise_dealer', 'franchise_make', 'front_legroom',
       'fuel_tank_volume', 'fuel_type', 'has_accidents', 'height',
       'highway_fuel_economy', 'horsepower', 'interior_color', 'isCab',
       'is_certified', 'is_cpo', 'is_new', 'is_coup', 'latitude', 'length',
       'last_sold_date', 'list_price', 'make_id', 'longitude',
       'main_picture_url', 'major_options', 'make_name', 'maximum_seating',
       'mileage', 'model_name', 'owner_count', 'power', 'price', 'salvage',
       'savings_amt', 'seller_rating', 'sp_id', 'sp_name', 'thft_title',
       'torque', 'transmission', 'transmission_display', 'trimId', 'trim_name',
       'wheel_system', 'wheel_system_display', 'wheelbase', 'width', 'year'],
       dtype='object')

fuel_mean=cars_df['city_fuel_economy'].mean()
fuel_mean
22.399261264511225

cars_df['city_fuel_economy'].fillna(value=fuel_mean, inplace=True)

cars_df['city_fuel_economy']=cars_df['city_fuel_economy'].replace('NaN',fuel_mean)

cars_df[ 'city_fuel_economy']

0           22.399261
1           22.399261
2          17.000000
3           22.399261
4           22.399261
...
1048570    14.000000
1048571    18.000000
1048572    22.399261
1048573    22.399261
1048574    15.000000
Name: city_fuel_economy, Length: 1048575, dtype: float64
```

*Replacing
Values*

Data Cleaning

- Null Values
- Replacing values

*Null
Values*

Replacing Values in columns

```
cars_df['fuel_type'].unique()
array(['G', 'D', nan, 'B', 'Flex Fuel Vehicle', 'E', 'H',
       'Compressed Natural Gas'], dtype=object)

cars_df['fuel_type']=cars_df['fuel_type'].replace(['G','D','B','E','H'],['Gasoline','Diesel','Biodiesel','Electric',
                                                       'Hybrid'])

cars_df['fuel_type'].unique()
array(['Gasoline', 'Diesel', nan, 'Biodiesel', 'Flex Fuel Vehicle',
       'Electric', 'Compressed Natural Gas'], dtype=object)

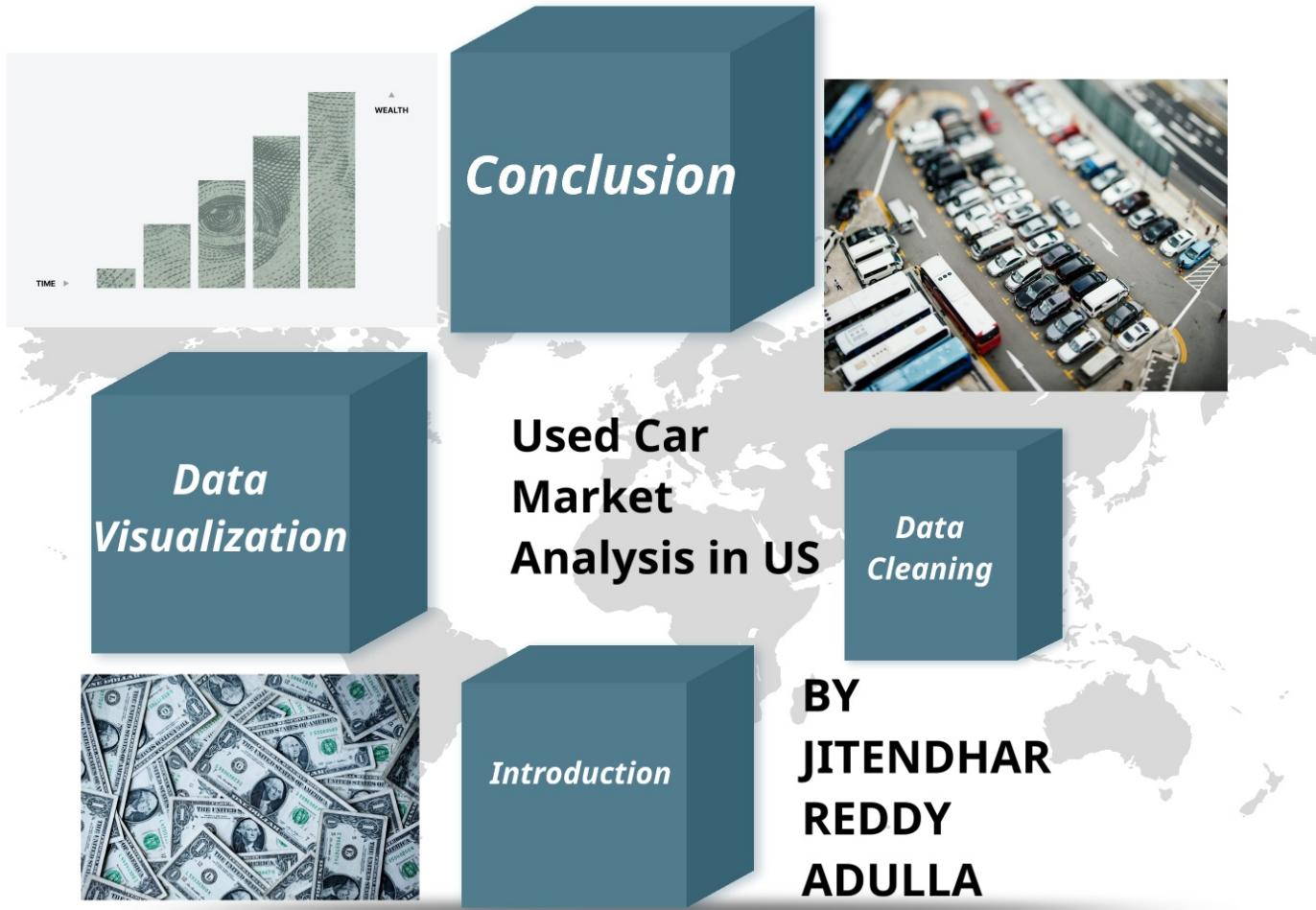
array(['Gasoline', 'Diesel', nan, 'Biodiesel', 'Flex Fuel Vehicle',
       'Electric', 'Hybrid', 'Compressed Natural Gas', 'Propane'],
      dtype=object)
```

*Replacing
Values*

Data Cleaning

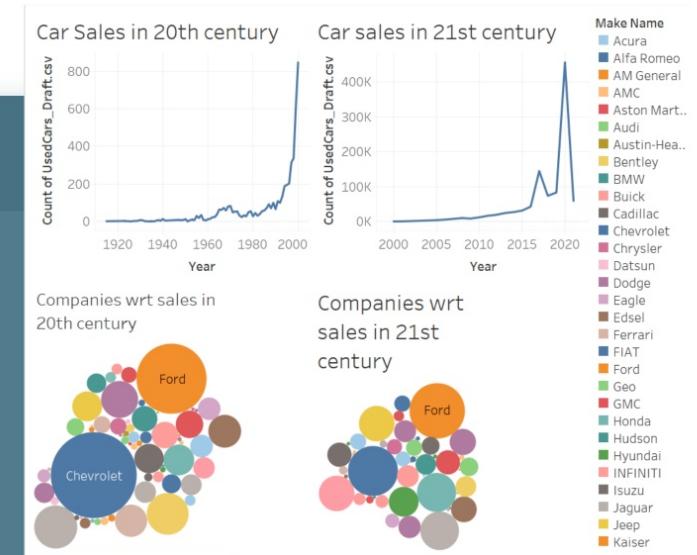
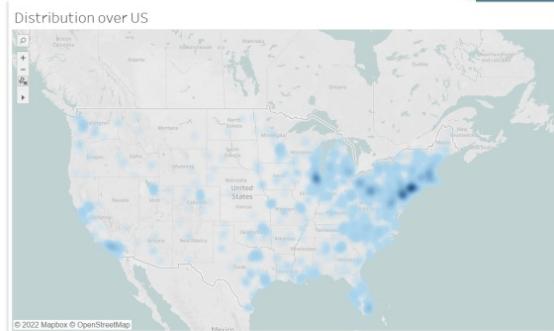
- Null Values
- Replacing values

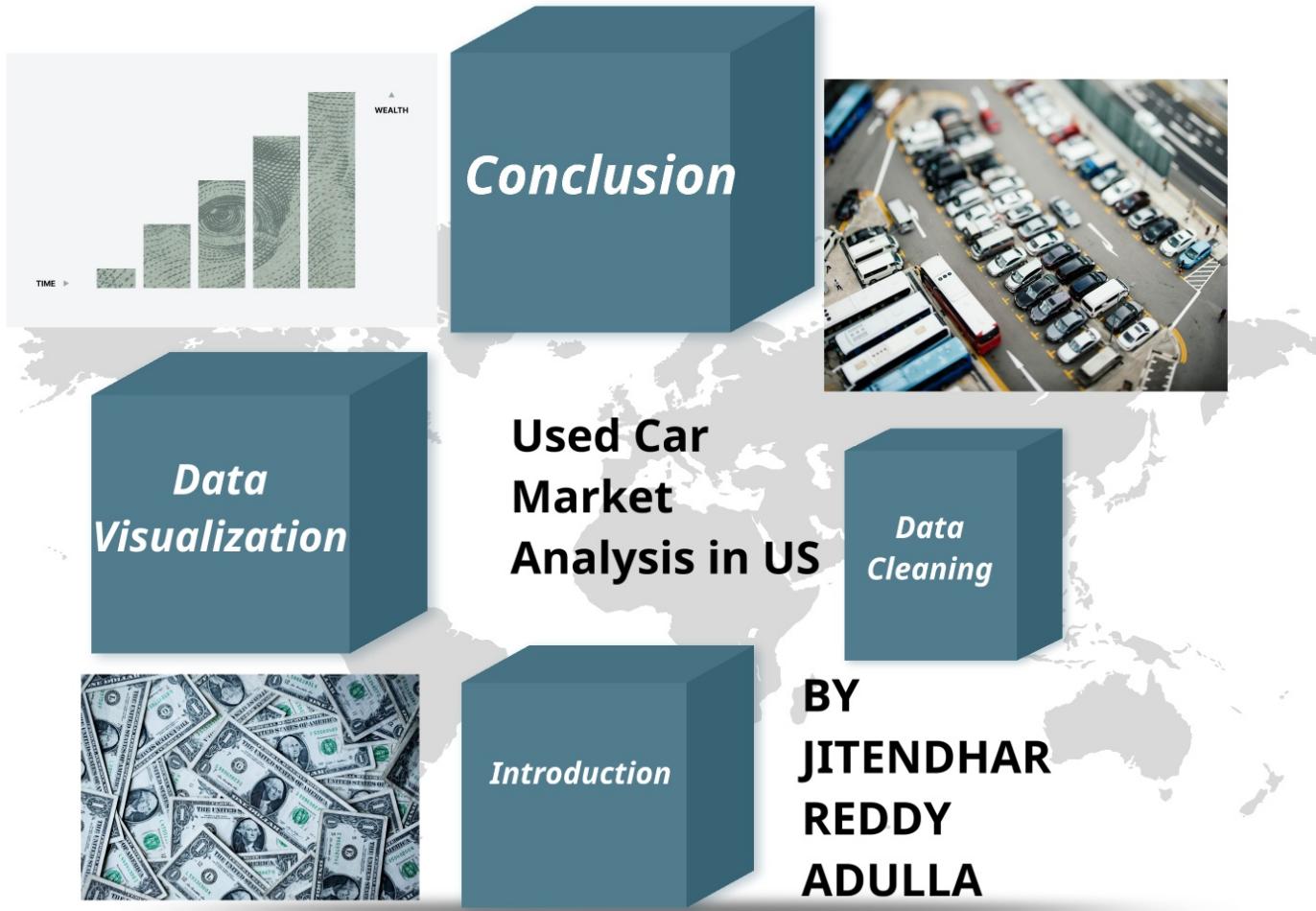
*Null
Values*



Data Visualization

- Visualization is performed using Tableau
- Here , are few visualizations .
- Lets look into the Tableau dashboard.





Conclusion

- Anlayzed the given Dataset.
- Factors effecting Price:
- Make Name
- Bodytype
- Color
- Fuel type

