**Capstone Two: Project Ideas**

Idea1: Cary, NC real estate price prediction

Build a model that more accurately predicts the real estate price sold at Cary, NC.

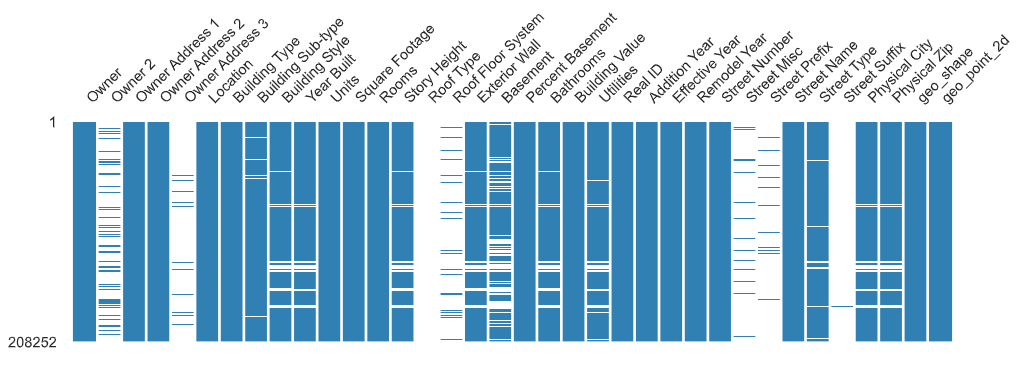
Question: Is the residential house in Cary, NC appropriately priced amid the highly competitive housing market? How to remodel a house to increase its value within a limited budget?

Description: Cary's real estate market is competitive, with an average of 5 offers per home and a swift sale time of around 37 days. To accurately assess the pricing of a house, it's important to not only consider the current market conditions but also determine its true cost and identify key features that can efficiently enhance its value.

Data source: Town of Cary government website open data. 1 csv file, 208252 rows, 36 columns.

The dataset has datetime variables, continuous numeric variables, categorical variables, location variables, text (natural language) variables. Can location variables be treated as natural language variables? The target is a continuous variable called ***Building Value***.

Missing values for each columns:



<https://data.townofcary.org/explore/dataset/building-points/information/?disjunctive.building_type&disjunctive.building_sub_type&disjunctive.bldgstyle&disjunctive.yearbuilt&disjunctive.storyheight&disjunctive.basement&disjunctive.utilities&sort=owner>

Idea2, data science salary prediction

Build a model that more accurately predicts data science salaries.

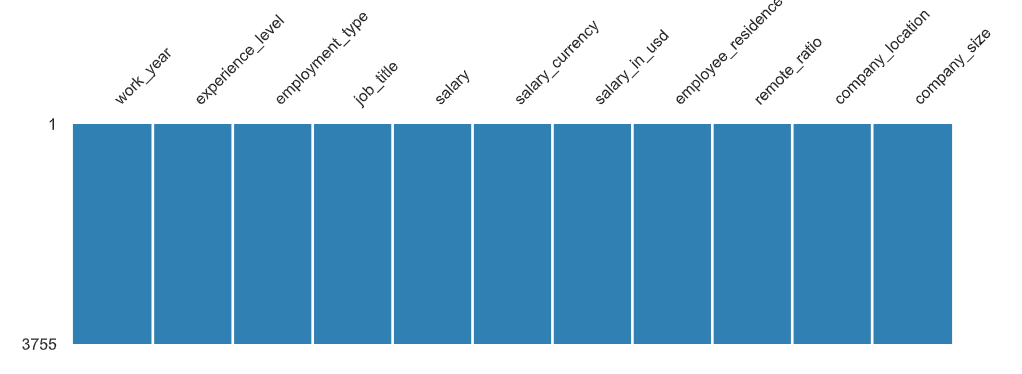
Question: Is Data Science a Good Career Choice? What is the predominant factor influencing the salary of a data science job in the United States? What’s the top 10 job titles in data science?

Description: With its escalating prominence and considerable impact, data science has evolved into a fascinating field for those deliberating their vocational journeys. Understanding the working and salary trends among the data science area worldwide can provide many useful insights to guide future studies.

Data source: Kaggle. 1 csv file, 3755 rows, 11 columns.

The dataset has datetime variables, continuous numeric variables, categorical variables, text (natural language) variables. The target is a continuous variable called ***salary\_in\_usd***.

Missing values for each columns:



<https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023>

Idea3: Cary, NC police incident hotspots prediction

Build a model that predicts the hotspots of police incidents at Cary, NC.

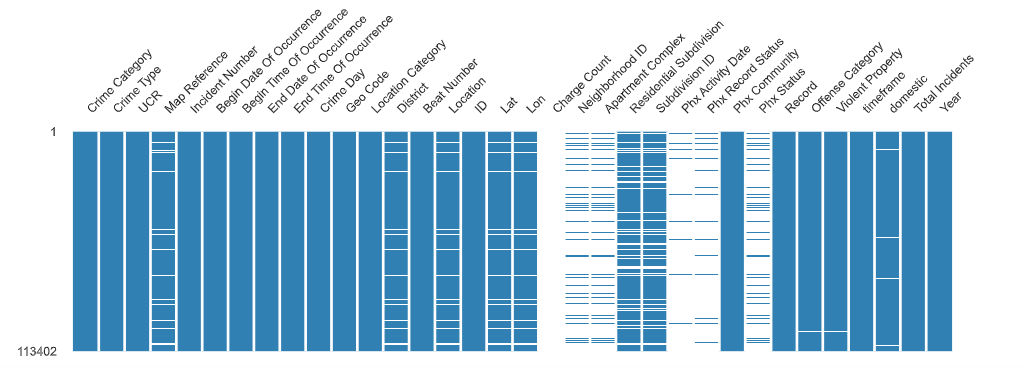
Question: What’s the most common police incident in a neighborhood? Where are the hotspots of police incidents at Cary?

Description: Crimes happen less often in Cary than in most of America. However, like any community, certain prevalent police incidents can still occur in the area. The most reported types of police incidents in Cary's neighborhoods typically include vandalism, fraud, simple assault, burglary, drugs and DUI.

Data source: Town of Cary government website open data. 1 csv file, 113402 rows, 34 columns.

The dataset has datetime variables, continuous numeric variables, categorical variables, location variables, text (natural language) variables.

Missing values for each columns:



[https://data.townofcary.org/explore/dataset/cpd-incidents/information/?disjunctive.crime\_category&disjunctive.crime\_type&disjunctive.crimeday&disjunctive.district&disjunctive.offensecategory&disjunctive.violentproperty&disjunctive.total\_incidents&disjunct](https://data.townofcary.org/explore/dataset/cpd-incidents/information/?disjunctive.crime_category&disjunctive.crime_type&disjunctive.crimeday&disjunctive.district&disjunctive.offensecategory&disjunctive.violentproperty&disjunctive.total_incidents&disjunctive.year)