POSITIVE INTEGER:PREDICTION ANALYSIS

PROBLEM STATEMENT

Companies pay commission to retailers to push their product for more sales. But at times the commission pay out wont be justified as the sales may not happen for the particular product.So our purpose is to propose a solution to identify when the commission pay out is not being justified and provide data visualizations through graphs . And provide our analysis of the dataset given.

ALGORITHMS USED:

* PRINCIPLE COMPONENT ANALYSIS (PCA):

1. **PCA** is an unsupervised statistical technique used to examine the interrelations among a set of variables. It is also known as a general factor analysis where regression determines a line of best fit.
2. They are used in data analysis and in machine learning for predictive models.

Mainly the PCA has two techniques

* Feature Elimination
* Feature Extraction

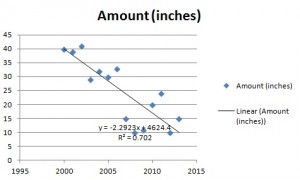
So we use feature elimination in which we reduce the feature space by eliminating features. Advantages of feature elimination methods include simplicity and maintaining interpretability of your variables. If we have a lot of data with the same form (meaning vectors of the same dimension...), your interest in this data is qualitative (you don't care the exact number itself, only the approximate number) and some of the data shows collinearity (dependency between vectors), PCA is a way to save storage space.

A close up of a map

Description automatically generated

* LINEAR REGRESSION :

Regression analysis is commonly used in research as it establishes that a correlation exists between variables. But [correlation is not the same as causation](https://www.thebalancesmb.com/quantitative-research-advantages-and-disadvantages-2296728). Even a line in a simple linear regression that fits the data points well may not say something definitive about a cause-and-effect relationship.In simple linear regression, each [observation](https://www.thebalancesmb.com/market-research-deductive-versus-inductive-2296727) consists of two values. One value is for the dependent variable and one value is for the independent variable. Simple Linear Regression Analysis The simplest form of a regression analysis uses on dependent variable and one independent variable.  For example,

FINAL RESULT

We noticed that the dataset was too big to directly involve any statistical algorithm directly. So we used PRINCIPLE COMPONENT ANALYSIS for data compression aka dimension reduction. Principle component analysis basically is used for feature extractions. The feature candidates are chosen based upon their variation .If the variation is too high that variable can be chosen as a feature. And then we are used linear regression by splitting the given data into testing and training data. Commission and unit\_sold would contribute the revenue. So we used linear regression to represent the above finding. Also we calculated the profit percentage by using the formula (commission/revenue)\*100. We then sorted the profit percentage in the descending order,so that we can get the product with maximum profit . We changed the months in terms of number and plotted a graph between the months and the profitability percentage to identify which month had the good profit.

A picture containing text

Description automatically generated

( Visualization of relationship between unit sold,revenue)

