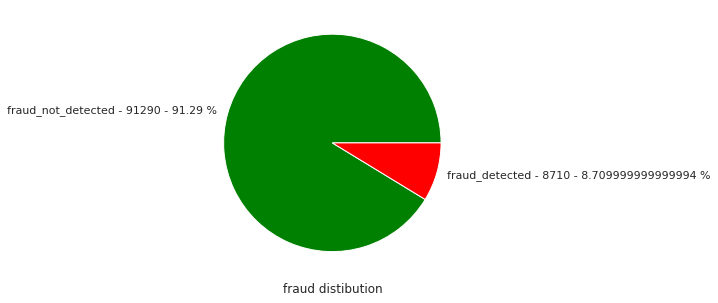
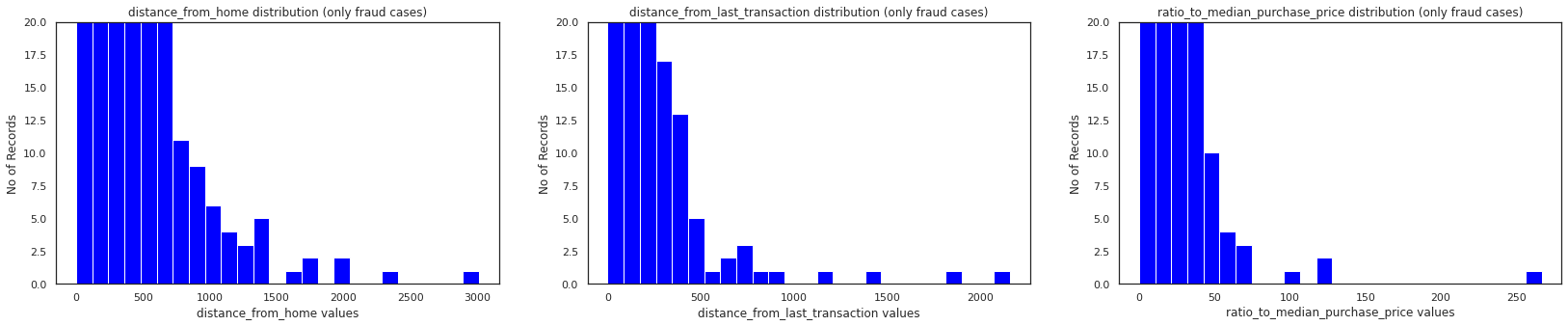
**Conclusion**

* **8710 records** are fraud detecting cases out of 1,00,000 records **(8.71%)**

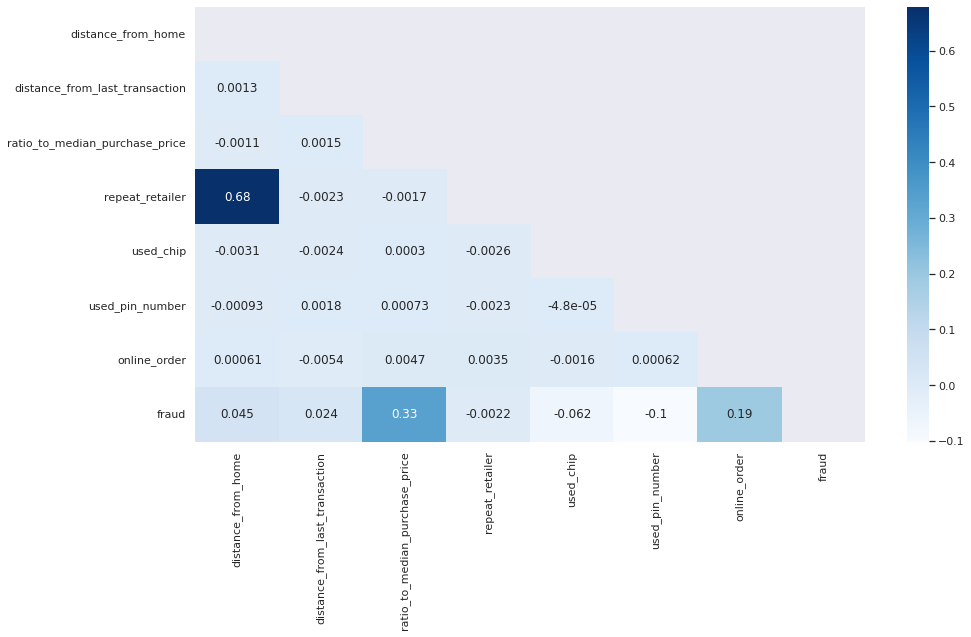


* Almost all fraud transactions are done by not using of pin numbers **(99.6%)**
* **6492 cases** of fraud transactions are not using of chip out of 8710 fraud transaction cases (**74.5%**)
* **94.5%** of fraud transactions are online\_ordered
* **7663** records of fraud transactions are caused by repeat\_retailer out of 8710 fraud transaction records (**88%**)



* Distribution of distance\_from\_home,distance\_from\_last\_transaction,ratio\_to\_median\_purchase\_price are **right skewed** which it may consists of some **outliers** above upper limit

**Conclusion in Correlations**



* The target variable is weakly correlated with all independent features
* The highest positive correlation value which the target variable having is with ratio\_to\_median\_purchase\_price(0.33) followed by online\_order(0.19)
* The weak negative correlation value which the target value having is with used\_pin\_number (-0.1) and very negligible negative correlation is with repeat\_retailer(-0.0022) & used\_chip(-0.062)
* distance\_from\_home & repeat\_retailer are having some strong correlation with each other**(multicollinearity)** but both are very negligibly correlated with target variable fraud
* distance\_from\_home,distance\_from\_last\_transaction,repeat\_retailer & used\_chip are negligibly correlated with target variable.So,it can be neglected while training the model