One of the most basic questions we might ask of a model is *What features have the biggest impact on predictions?*

*feature importance*

 - *supervised regression*machine learning task

* Given pickup and dropoff locations, the pickup timestamp, and the passenger count, the objective is to predict the fare of the taxi ride.
* Fare\_amount – Target
* Models to use: linear regression, random forest
* The reason a [random forest](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm" \t "_blank) typically outperforms a linear regression is because it has more flexibility — lower bias — and it has reduced variance because it combines together the predictions of many decision trees. A linear regression is a simple method and as such has a high bias — it assumes the data is linear. A linear regression can also be highly [influenced by outliers](https://stats.stackexchange.com/questions/350775/influential-observations-and-outliers-in-linear-regression-model" \t "_blank)because it solves for the fit with the lowest sum of squared errors.







