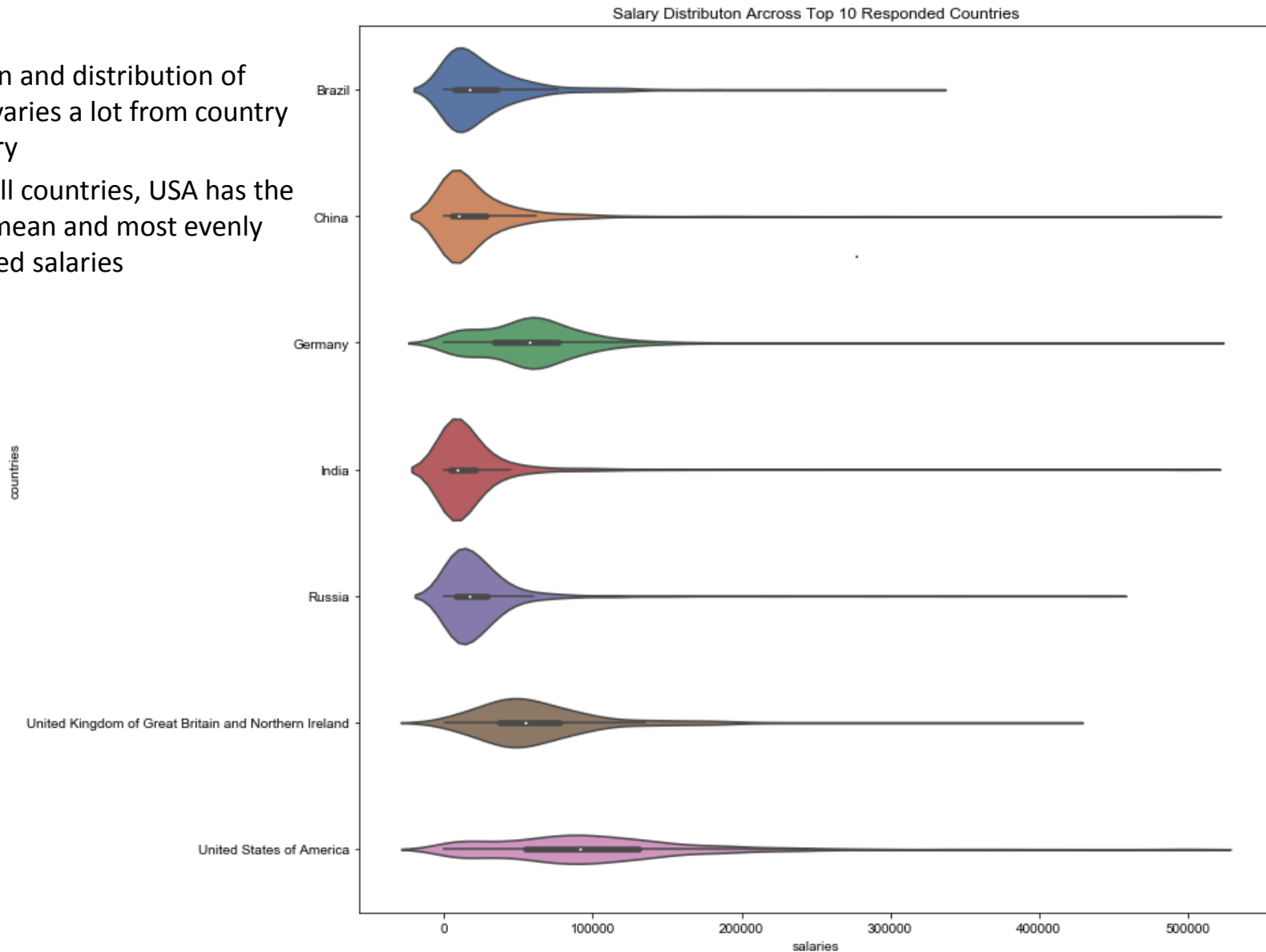


Exploratory analysis

-Country

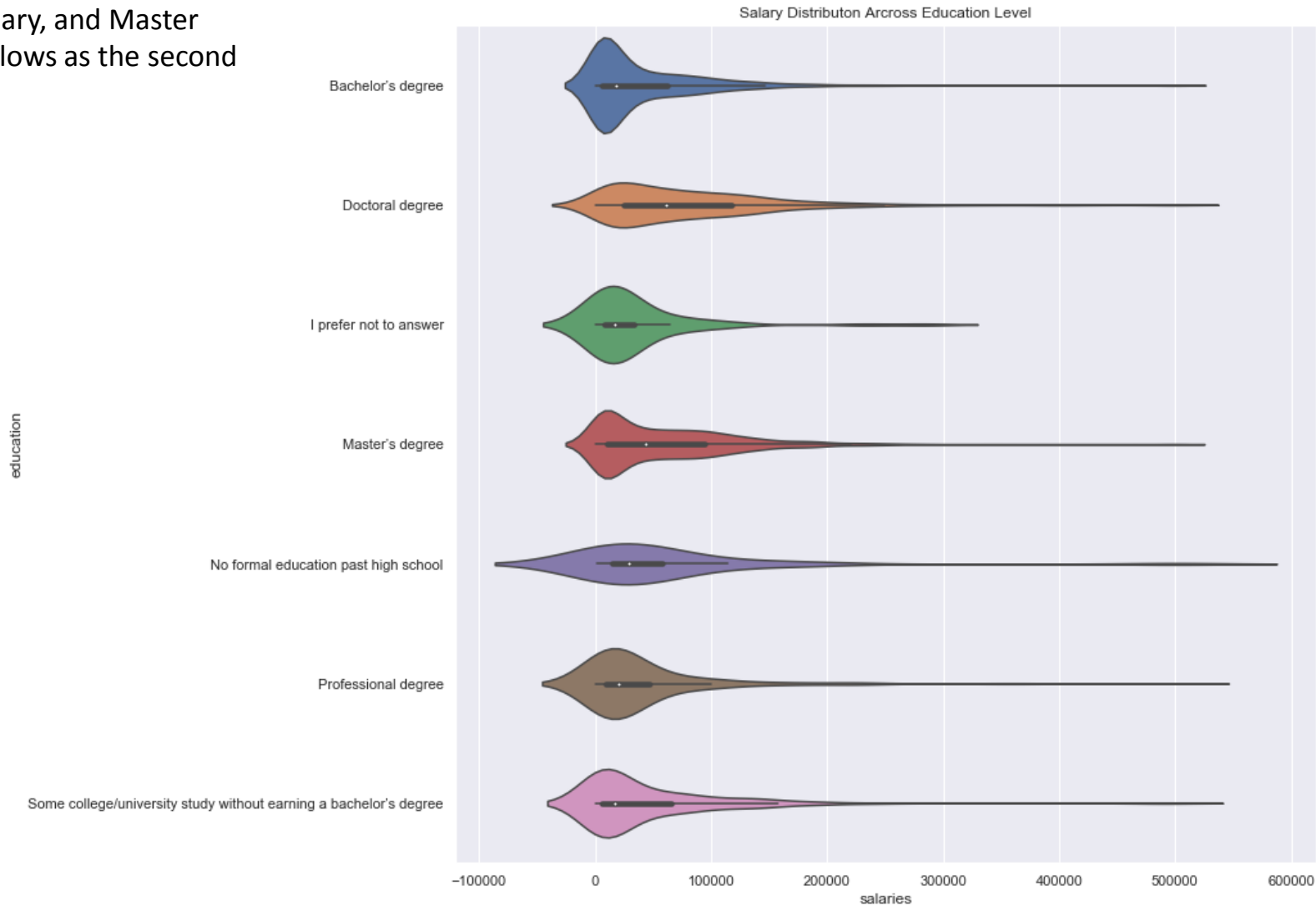
- The mean and distribution of salaries varies a lot from country to country
- Among all countries, USA has the highest mean and most evenly distributed salaries



Exploratory analysis

-Education Level

- Income varies in different level of education
- As expected, PHD earns the biggest salary, and Master Degree follows as the second

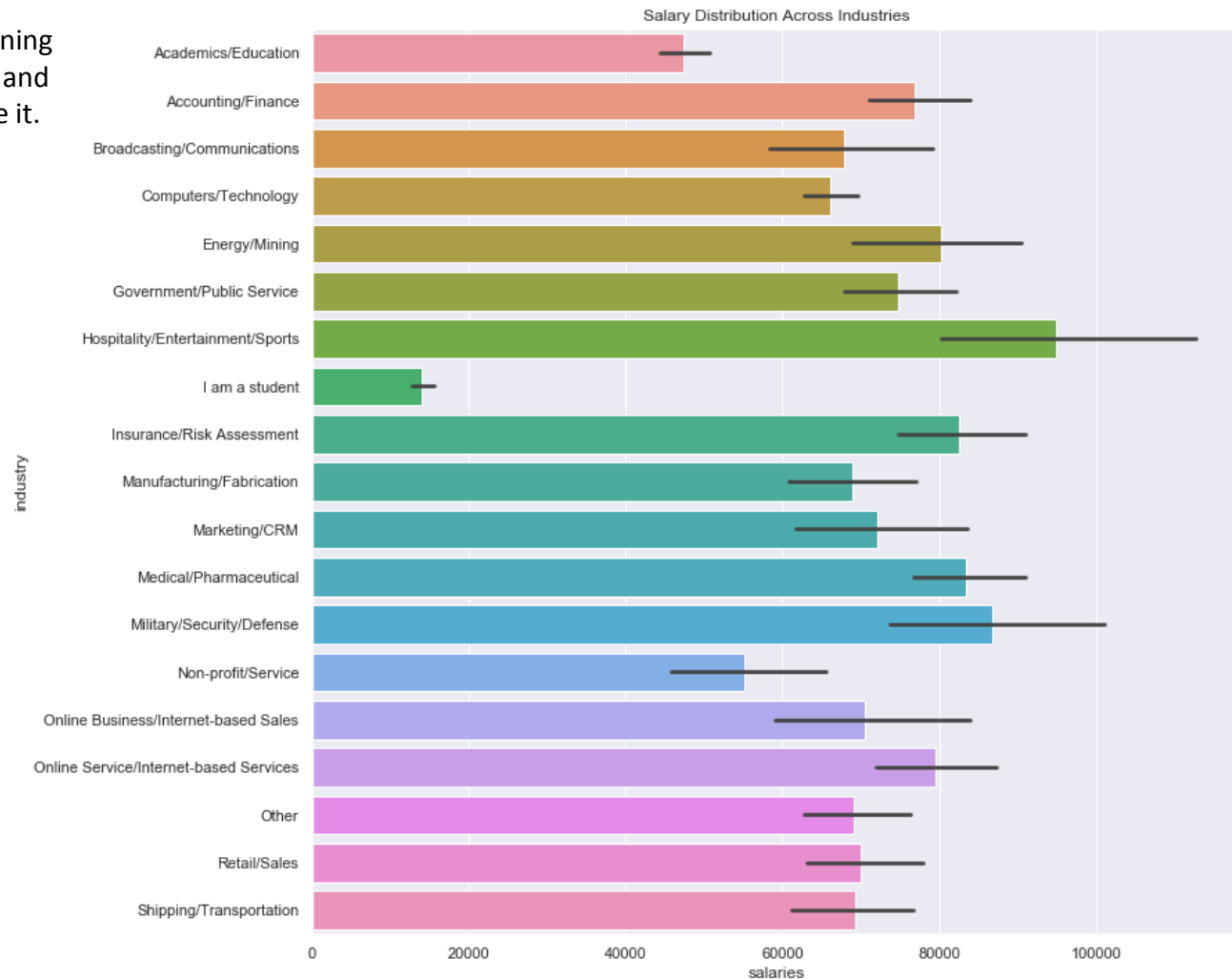


Exploratory analysis

-Industry

The top earning industry is Hospitality/
Entertainment/ Sports! Accounting /Finance
is not as high as expected.

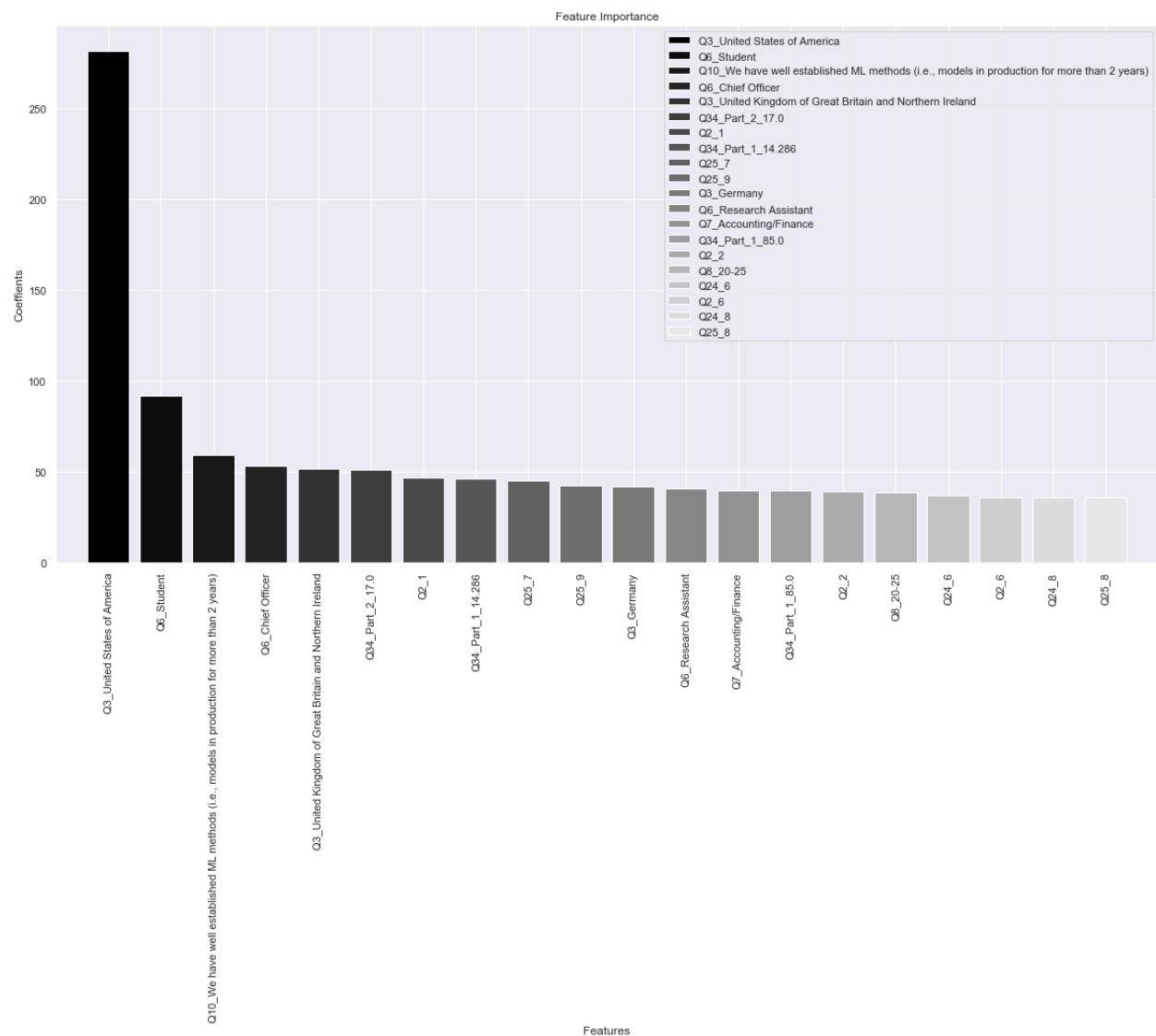
Military/Security/Defense, Energy/Mining
Online Business/Internet-based Sales and
Medical/Pharmaceutical are all before it.



Feature Importance

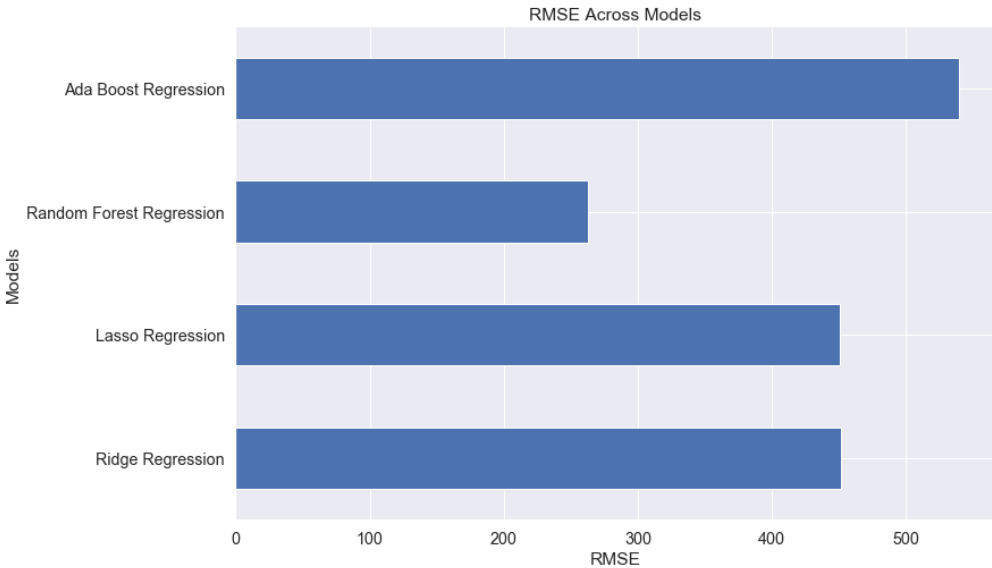
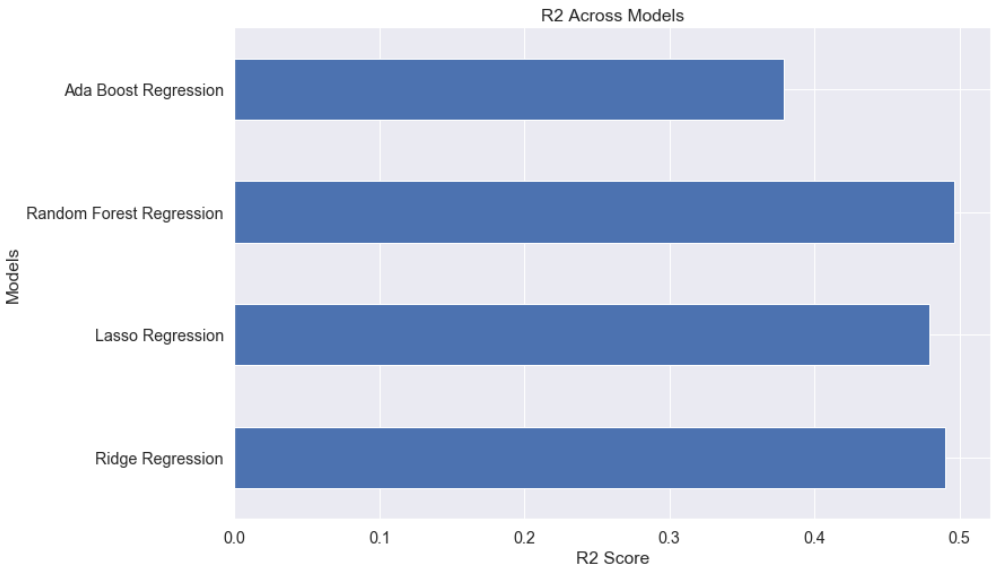
Important features are:

1. Country
2. Age
3. Use of well established ML models in the company (Q10)
4. Title of position (of course)
5. How many time devoted to data gathering and cleaning(Q34) (the importance of data engineering!)



Optimal Model Choosing

As the grid search result shows, Random Forest Regression perform the best. It has the highest R2 score at 0.496 and lowest RMSE at 263



	Ridge Regression	Lasso Regression	Random Forest Regression	Ada Boost Regression
R2	0.490263	0.479087	0.496417	0.378912
RMSE	451.442589	450.979869	262.608038	539.544114