Python:

1) **Sklearn**- For Machine Learning algorithms. This library has almost all the important machine learning algorithms used for industries.

2) **Pandas, Matlotlib & Seaborn-** Mostly used together for analytics and visualisation work.

3) **TensorFlow and Keras-** Used for Deep Learning.

4) **Numpy and Scipy-** Used for scientific computations.

5) **nltk-** Used for Natural Language Processing.

R:

1) **dplyr-** dplyr is the package for fast data manipulation.

2) **ggplot2 and shiny-** R's famous package for making beautiful graphics. Python visuals are nowhere near to the quality of visuals created using this library.

3) **caret and randomforest-** For Machine Learning purpose.

4) **tidyr-** Tools for changing the layout of your data sets.

5) **stringr-** Easy to learn tools for regular expressions and character strings.

**Leaflet (folium in Python) and Plotly are common libraries in both languages, and are used to create interactive plots like geomaps, etc.**

Conclusions

Some brief insights that we gathered from the notebook:

1) Majority of the respondents are from USA followed by India. USA also had the maximum number of data scientists followed by India. Also the median Salary is highest in USA.

2) Majority of the respondents are in the age bracket 20-35, which shows that data science is quite famous in the youngsters.

3) The respondents are not just limited to Computer Science major, but also from majors like Statistics, health sciences,etc showing that Data Science is an interdisciplinary domain.

4) Majority of the respondents are fully employed.

5) Kaggle, Online Courses(Coursera,eDx,etc), Projects and Blogs(KDNuggets,AnalyticsVidya,etc) are the top resources/platforms for learning Data Science.

6) Kaggle has the highest share for data acquisition whereas Github has the highest share for code sharing.

7) Data Scientists have the highest Job Satisfaction level and the second highest median salary (after Operations Research Analyst). On the contrary, Programmers have the least Job Satisfaction level and one of the least median salary also.

8) Data Scientists also get a hike of about 6-20% from their previous jobs.

Tips For Budding Data Scientists

1) Learn **Python,R and SQL** as they are the most used languages by the Data Scientists. Python and R will help in analytics and predictive modeling while SQL is best for querying the databases.

2) Learn Machine Learning Techniques like **Logistic Regression, Decision Trees, Support Vector Machines**, etc as they are most commonly used Machine Learning techniques/algorithms.

3) **Deep Learning and Neural Nets** will be the most sought after techniques in the future, thus a good knowledge in them will be very helpful.

4) Develop skills for **Gathering Data** and **Cleaning The Data** as they are the most time consuming processes in the workflow of a data scientist.

5) **Visualisations** are very important in Data Science projects and almost all projects require Visualisations for understanding the data better. So one should learn Data Visualisation as Data Scientists consider it to be a **necessary or nice to have skill.**

6) **Maths and Stats** are very important in Data Science, so we should have good understanding of it for actually understanding how the algorithm works.

7) **Projects** are the best way to learn Data Science according to Data Scientists.So working on projects will help you learn data science better.

8) **Experience with ML Projects in company and Kaggle Competitions** are the best ways to show your working knowledge in Data Science. Working on ML projects in a company gives the experience of working with real world datasets, thereby enhancing the knowledge. Kaggle competitions are also a great medium, as you will be competing with Data Scientists over the world. Also a **Kaggle Rank** can be a good USP in the resume.