**STATISTICS PROJECT**

**TO**

**FIND OUT REASONS FOR HIGH DATA SCIENCE JOB DEMAND AND REASONS FOR SWITCHES/CAREER PROGRESSION**

**TO DATA SCIENCE**

**ALLY SAHA-C22002**

**ANTARLIN CHANDA- C22003**

**DS JAN22**

**PROJECT BACKDROP AND MOTIVATION BEHIND THE PROJECT:**

The aim of this project study is to explore the distribution of variables and responses through Exploratory Data Analysis and to gain insights into the causes and probable factors of the switch or career transition of individuals to the booming field of Data Science and their expectations out of it.

Keeping ourselves in regular touch with blogs, and articles are written in the field of Data Science, we came upon an article that intrigued us with various interesting questions on the journey of an individual in the field of Data Science like which age-group individuals transition the most? Sectors in which individuals were involved before they entered this world. We also pondered upon the fact can we draw some inferences from such data about each individual who is switching his/her career. So this thought brought us to this project wherein we have tried to find out certain interesting inferences about switching people to Data Science by doing a primary survey keeping students from both the Kolkata and Bangalore campuses of Praxis Business School as our sample dataset. We attach below the blog article which intrigued our curiosity about this particular topic.

<https://www.kaggle.com/code/foolofatook/journey-into-the-data-science-industry>

Praxis launched its first full-time Business Analytics course which was later rechristened as Data Science. This course was offered full-time and with in-campus mode but ever since Covid hit in 2020, the course shifted to online mode and now it is being offered in dual-learning mode i.e., both offline and online. Praxis being a leading institution in helping students switch their careers to Data Science, would help us get the best insight for our aim. Our study aims to delve into the opinions and reviews of the students of Praxis and find out the probable causes of their career switch and where they see themselves settling on financial and professional grounds.

**INTRODUCTION:**

Data Science Definition

Data science combines multiple fields, including statistics, scientific methods, [artificial intelligence](https://www.oracle.com/in/artificial-intelligence/) (AI), and data analysis, to extract value from data. Those who practice data science are called data scientists, and they combine a range of skills to analyze data collected from the web, smartphones, customers, sensors, and other sources to derive actionable insights.

Data science encompasses preparing data for analysis, including cleansing, aggregating, and manipulating the data to perform advanced data analysis. Analytic applications and data scientists can then review the results to uncover patterns and enable business leaders to draw informed insights.

Dataset description

The dataset contains responses from 50 students from Praxis Business School’s  PGPDS course, which contains opinions on a varied number of parameters related to  their previous job sectors and what do they expect from the field of Data Science.

Below, we’ll explore each of the variables in the dataset one by one and try to have an extended understanding and distribution of it across other variable features using EDA.

**DATA COLLECTION METHODOLOGY**

Due to the requirements of the study, We did a primary survey to collect our Data. We made a Google Form Questionnaire comprising 21 Questions to find various attributes of our data points.

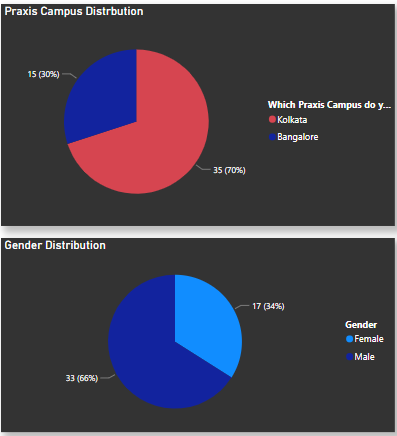
We sent the Form to Our Kolkata and Bangalore batchmates.

Since most people in the DS course are from a wide variety of fields and industries and are trying to get into the DS Domain, so we chose this dataset as we would get a very wide range of data that we needed to do a study like this.

Questions asked were Which College Campus, Age, Gender, Resident State, Graduation Year, Graduation Field and Subject, PG if any, Coding and Mathematics, Statistics background, knowledge source about DS Field, Job Experience, Work Experience Duration, Industrial Sector, and Job Role/Domain, Previous Salary Drawn, Career Transition Reason, Expected Salary, Desired DS Job Profile, Preferred Work City.

Distribution of responses

The survey generated responses from 50 students from different batches of the data science course offered by Praxis Business school. It also incorporates both the campuses of Praxis Business School i.e., Bangalore and Kolkata.



Insights –

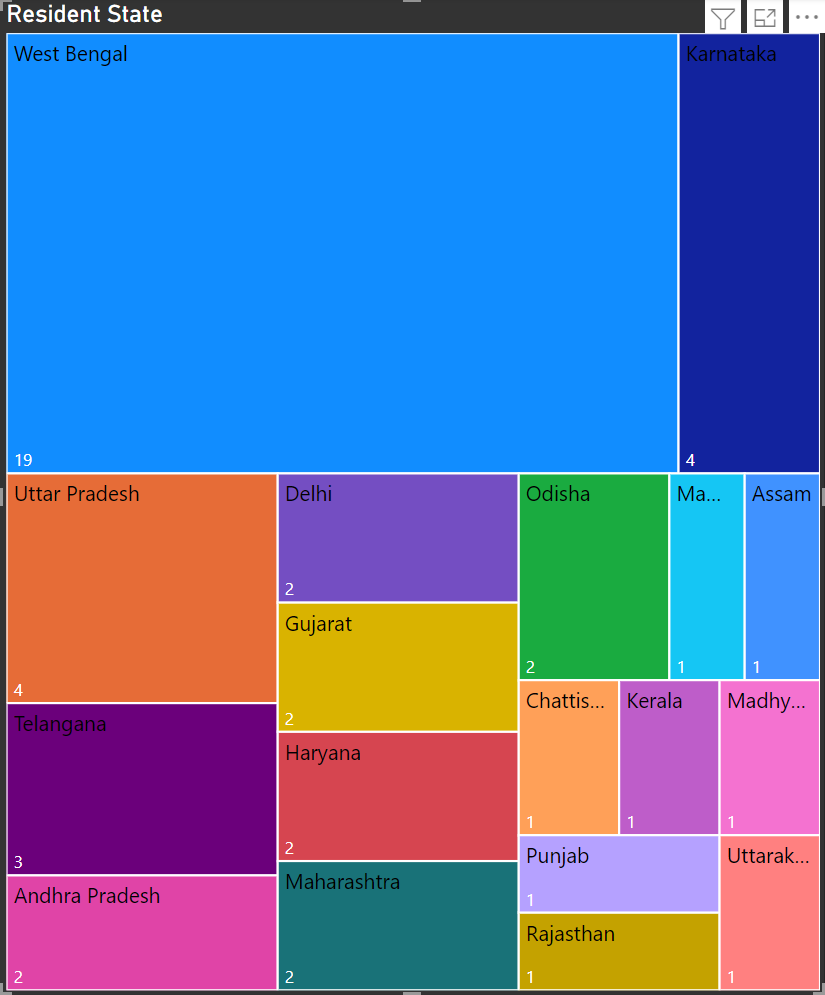
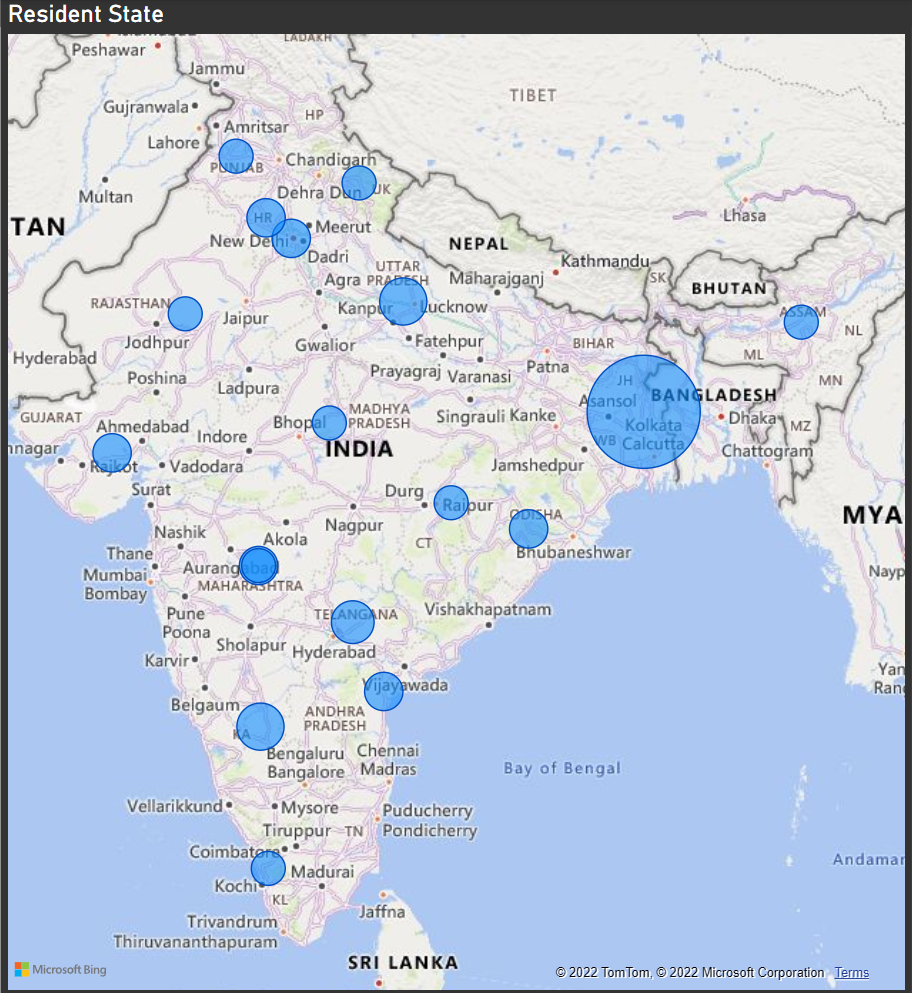
• From the table, we can see that we received 15 responses for Bangalore  campus and 35 responses for Kolkata campus.

• For both the campus locations, majority of responses came from the newly  started batch i.e. Jan 2021 which amounts to 30% for Bangalore campus and  70% for Kolkata campus.

* From the chart we can clearly see that 66% of the batch comprises males whereas only 34% of the batch comprises females.

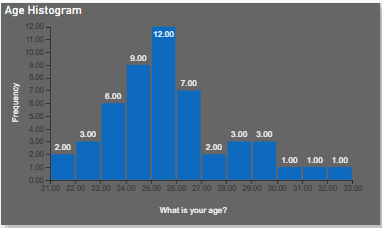
**Exploratory Data Analysis**

**Feature Analysis**



We have used Country Chart and Tree Map to showcase the distribution of home residence of students in our sample dataset over various states in our country.

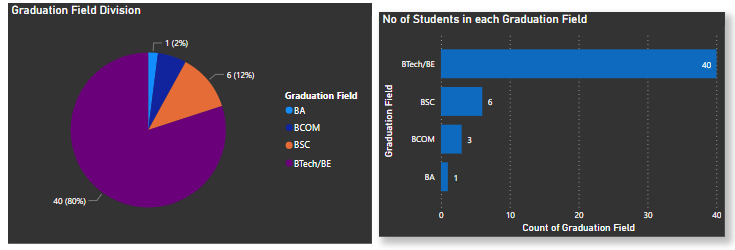
As we can see from both the charts, maximum students from our sample dataset belong to West Bengal.Nearly 38% of students are from West Bengal;8% are from Karnataka;Uttar Pradesh too accounts for 8% of students.Telangana and Maharashtra account for 6% of students each.The states of Gujarat,Haryana,Odisha and Delhi account for 4% of students each.The states of Punjab,Assam,Kerala ,Rajasthan,Andhra Pradesh,Madhya Pradesh and Uttarakhand account for 4% of students respectively.

Mean = 25.74 yrs

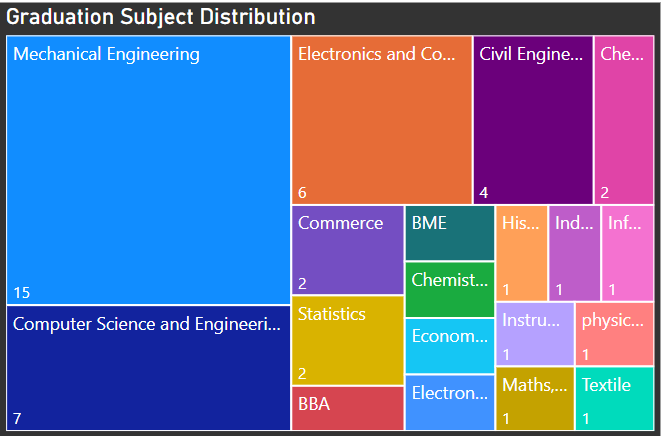
Median = 25.41 yrs

Mode = 25.37 yrs

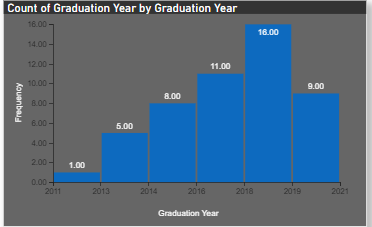
We wanted to represent which age group individuals in our dataset transition the most to Data Science or rather enter into this field quitting their previous fields.As we can see individuals of age range from 25-26 yrs transition the most. Freshers in the age range of 21-22 yrs and also individuals from 30 onwards show lesser transition than the rest age groups.Freshers probably get diverted into a lot different sectors and individuals above 30 get settled in their jobs and don’t switch easily.The mean being slightly more than the median shows that the graph is slightly right skewed.



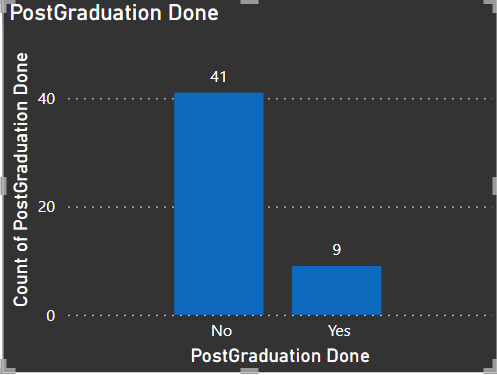
The above two graphs clearly indicate that majority of students have an engineering background with 80% of the students(40 students) having done BTech/BE in their graduation followed by 12% of students(6 students) having done BSc and 6% and 2% of students from BCom and BA respectively.BTech students, already having done an application oriented degree which requires critical and logical thinking and mathematical application(all of which are also critical aspect of Data Science) probably find it easier to make this switch hence we get the above trends.



This Tree Map shows most of the transitions are made from Mechanical Engineering department probably due to less or restrictive growth opportunities in core engineering sector.

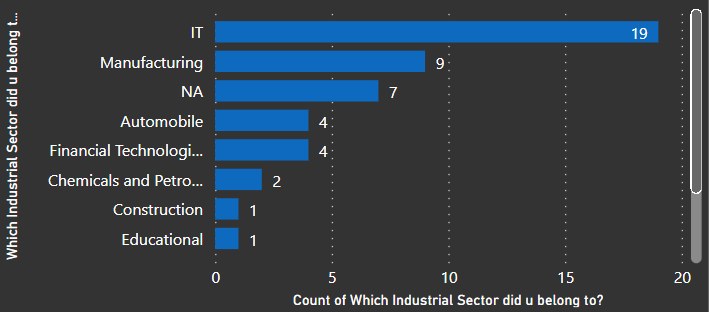


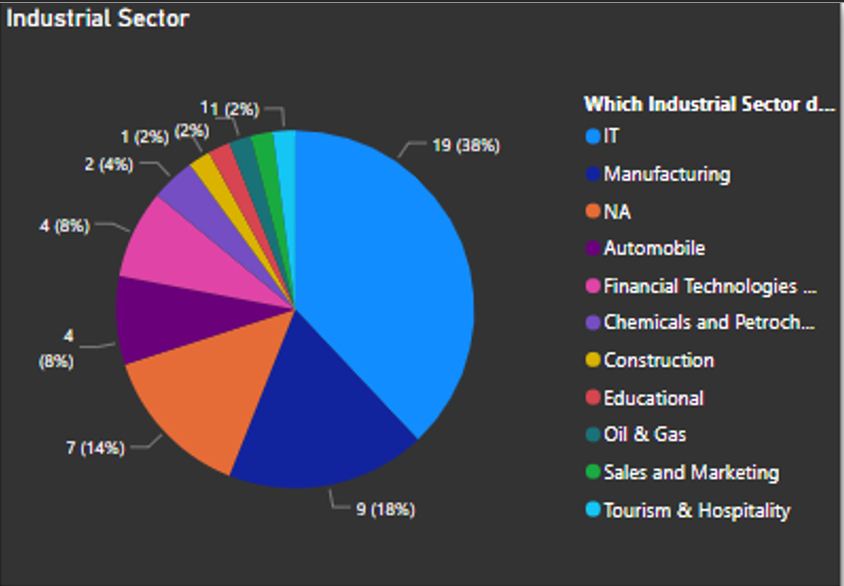
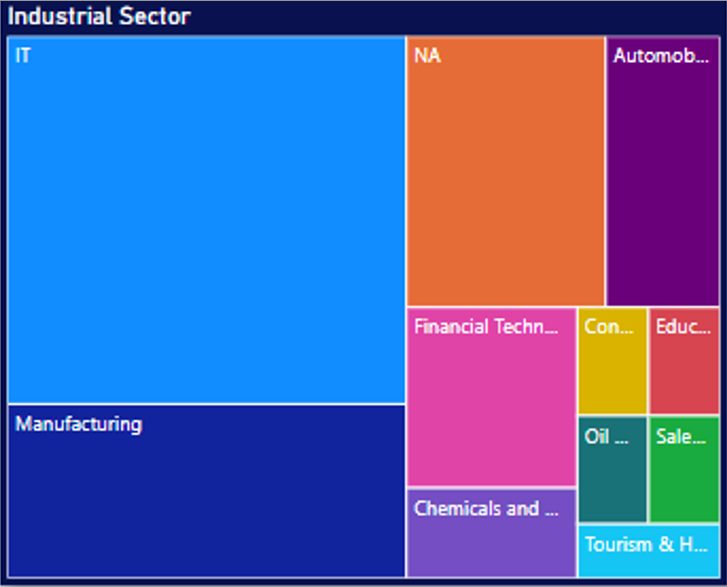
This graph shows the distribution of students according to their graduation year showing people graduating between 2018-2019 making maximum transitions.



Majority of students in our sample dataset have not done any Post-Graduation course previously.Nearly 40 of the students have enrolled into PGPDS course with

just Bachelors Degree.

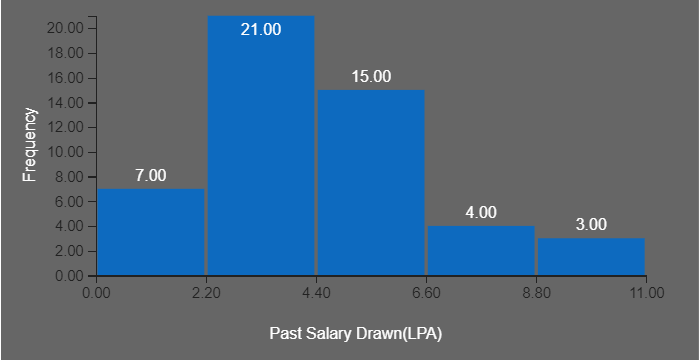




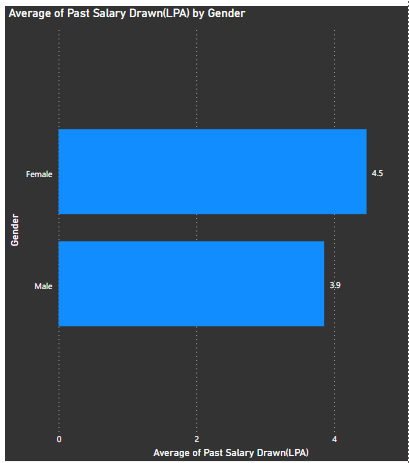
This chart represents a very interesting aspect of career transition. We can clearly see IT professionals are the largest set of individuals who are making transition to Data Science probably due to various common aspects such as sound knowledge of computers and similar working environment.



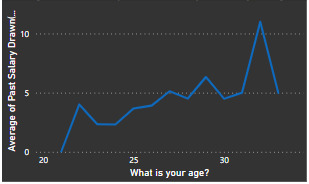
This bar graph represents distribution of individuals according to number of working years.As we can clearly see most of the people switch after they have got 3 years of work experience.We also find good number of people transitioning with 1-2 years of experience.Generally people stop switching after they have gained considerable experience in one field , as a result of which we see less number of people transitioning with greater number of work experience.



This piece of analysis clearly shows us that individuals with 2-4 lpa of annual income make most transitions probably because they want to increase their annual income in a booming field like Data Science.Individuals with 8-11 lpa of annual income are probably satisfied with their annual incomes ,hence transitions in such income range are lot lesser.

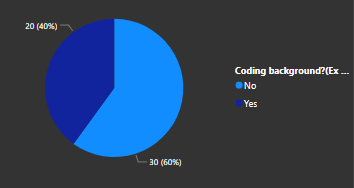


We tried to find out average past salary of the students.Females in our dataset used to get salaries ranging from 4-5.5 lpa owing to which they have an average salary around 4.5 whereas average salary of males is coming around 3.9 lpa .However average salaries of both males and female spans around 4-4.5 lpa before switching to this field.



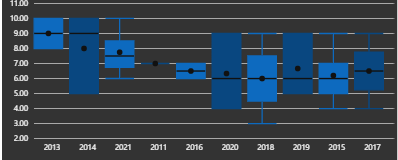
We also tried to find out a trend between average past salary and age of individuals.We can clearly see individuals with age above 30 yrs getting more salary than other age groups on an average probably owing to more work experience.

**Course Related Analysis**

****

We wanted to know how many students had a previous coding background before they decided to switch and we found out nearly 60% come from a non-coding background.

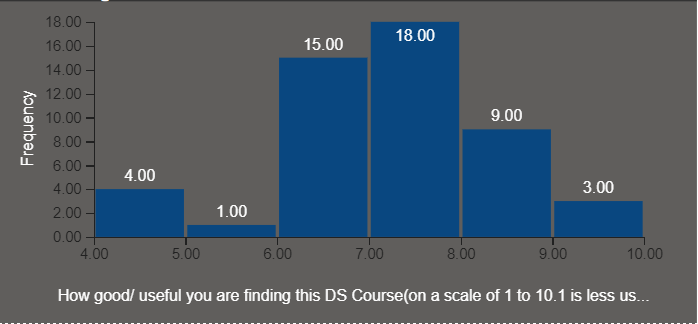


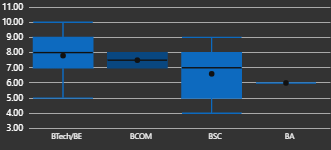
****

Transitioning career is not an easy job and especially to a highly logical mathematical oriented field might lead to finding this field difficult for some individuals.Keeping this mind we asked our participants to rate the difficulty of the course according to their current perspective on the basis of 1-10 , 1 being the easiest to 10 being most difficult and majority of students rated it between 4-6 which shows mostly people are finding this course of average difficulty level.

Mean difficulty level was found to be 6.63 and median 6.

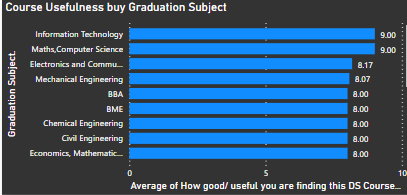
Also we noticed from the box plot an interesting trend of people whose graduation year is earlier(and thus age higher)are on average finding the course more difficult than the younger people.This we attributed to the fact than younger people may have more grasping power for newer information than older people.



****

We were also inquisitive about the fact that whether students are actually finding the PGPDS course in Praxis useful or not.So we asked them to rate the course on a scale of 1-10,1 being the least useful and 10 being the most useful.Since we got most of the answers with a rating of 7-8 , with a mean of.Mean usefulness rating was found out to be 7.7 and median 8 **.**This can be inferred as students find this course fairly good in terms of usefulness.

Also from box plot we found out people from BTech/BE are finding the course more useful as medians are higher so is mean. This is possibly due to the fact that for many tech related fields DS is a natural career progression.

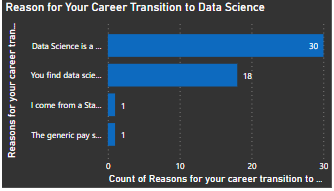


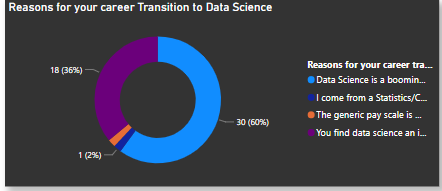
From this graph we can clearly see students from varied educational backgrounds find the course almost equally useful in their step towards transition.



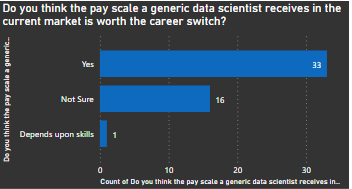
From this graph we can infer that students from various professional backgrounds too find the course also find it almost equally useful which shows the course is well-suited for people coming from various professional backgrounds.

**Reasons found out**



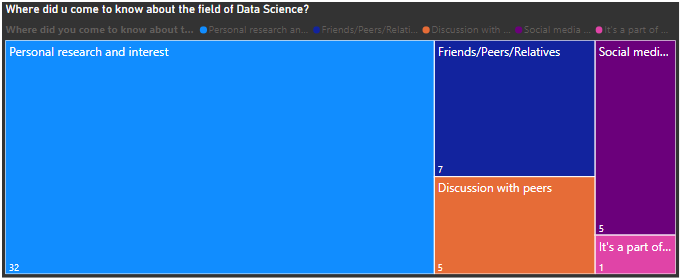


Transitioning in career is a big move which an individual makes after giving a lot of thoughts.So we wanted to know the major reason behind such a transition.We gave our participants reasons like – “Data Science being a booming field can give immense growth opportunities”, “They find Data Science an intriguing field and out of personal interest opted for it”,coming from a statistical background and generic higher pay scale were other reasons which we gave them.Majority of individuals have switched because Data Science is a booming field and can give them immense growth oppertunities.



We tried to estimate how many people found Data Science a financially lucrative field to work for?

We found about 66 percent felt DS is a highly paying field while abt 32 percent were unsure. Remuneration is a major factor when making a decision about a career switch as was found out.

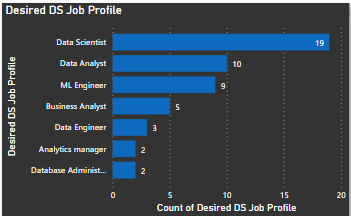


People become interested in something after they gain some interest or information towards it. Our participants were surveyed according to how they came across the DS field and what motivated them to pursue it. As clearly seen from treemap more than 63 percent of respondents pursued DS out of their own research and interest. About 14 percent from friends/peers and 10 percent each for people who discussed with peers and on social media platforms. We found this significant due to the fact that most career transitioning moves done by working professionals who have some degree of maturity and experience are independent and capable enough to make their decision and thus as a result personal research and interest was the one which is bringing most people in DS.



We get a very interesting trend from the above graph. Generally, people who have switched with greater number of work experience have switched because they have found that Data Science is a booming field and because of their personal interest in the field. On the other hand, people who have switched with lesser number of work experience have either done it due to having a common subject related to Data Science such as statistics or Computer Science or because of a generic higher pay scale than the previous job.

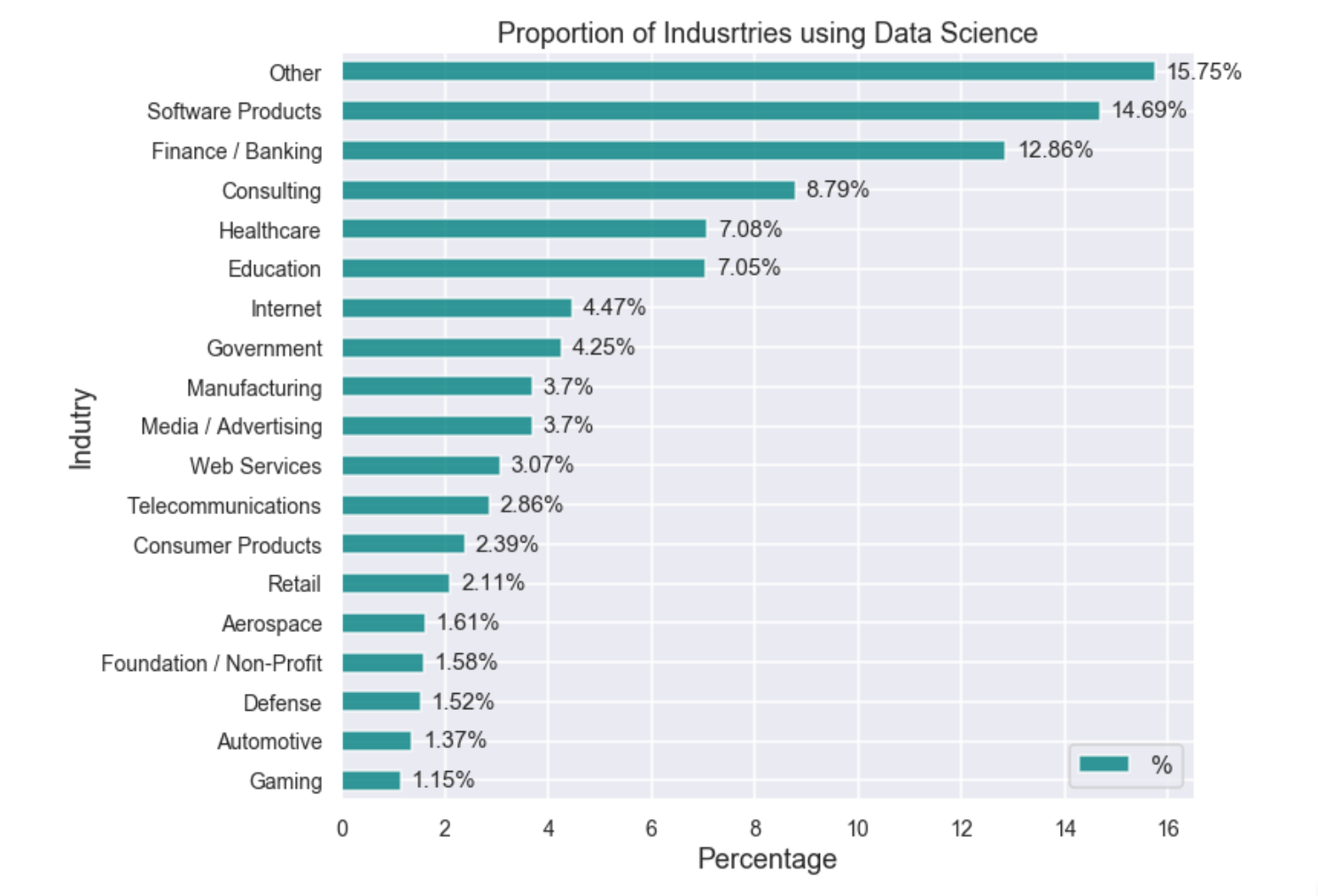
**Career Goals**

****

Most Desired Job Profile was found to be the Data Scientist followed by Data Analyst, ML Engineer, Business Analyst, Data engineer.

So many different positions showed us the breadth of data science jobs and roles available. For example

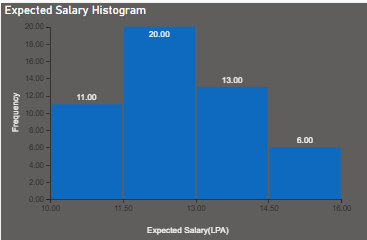
Data scientists being the most generic of all the roles need to know Python, SQL, Statistics, and Machine Learning all to a reasonably high degree, whereas Data Analyst roles require mostly Python, SQL, and BI Tools.ML Engineers are highly focused on AI and ML.DBA and Data Engineers are focused on RDBMS and NoSQL-related skills.



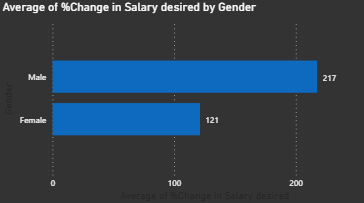
Note(above graph is external source)

Also Industries using Data Science span almost every sector imaginable. From IT, Banking, Fintech, Consulting, Education to Manufacturing, Retail, Defense, and telecommunications. By ‘Other’, it could mean various industries like — Research of different kinds, Medicines, Pharmaceuticals, E-commerce, Construction, Transportation, Insurance, Travel & Hospitality, Utilities, Natural resources & Energy, etc.

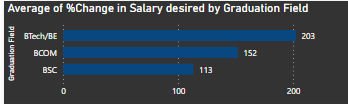
Keeping in mind the various nature of our input people profile we can easily see that no matter what sector they belonged to everyone can apply their domain knowledge in the field of Data Science and brings something to the table.



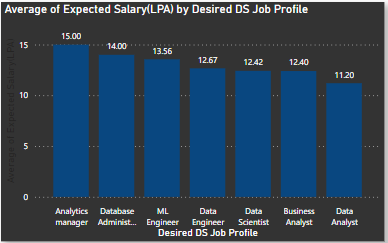
Salary Expectations of most people were found to range from 10-16lpa plus with the median lying in and around 12lpa range. This is inline with the fact that DS Jobs require high amount of skills and provide good remuneration.



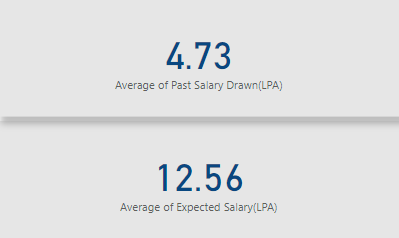
We find out that males are demanding a salary increase of almost triple than their previous while women demanding around 2 times more. This may be due to the fact that previously we found out male average previous salary(3.9 lpa) was less compared to females (4.5 lpa) and thus to come to the same level they are thus demanding more.



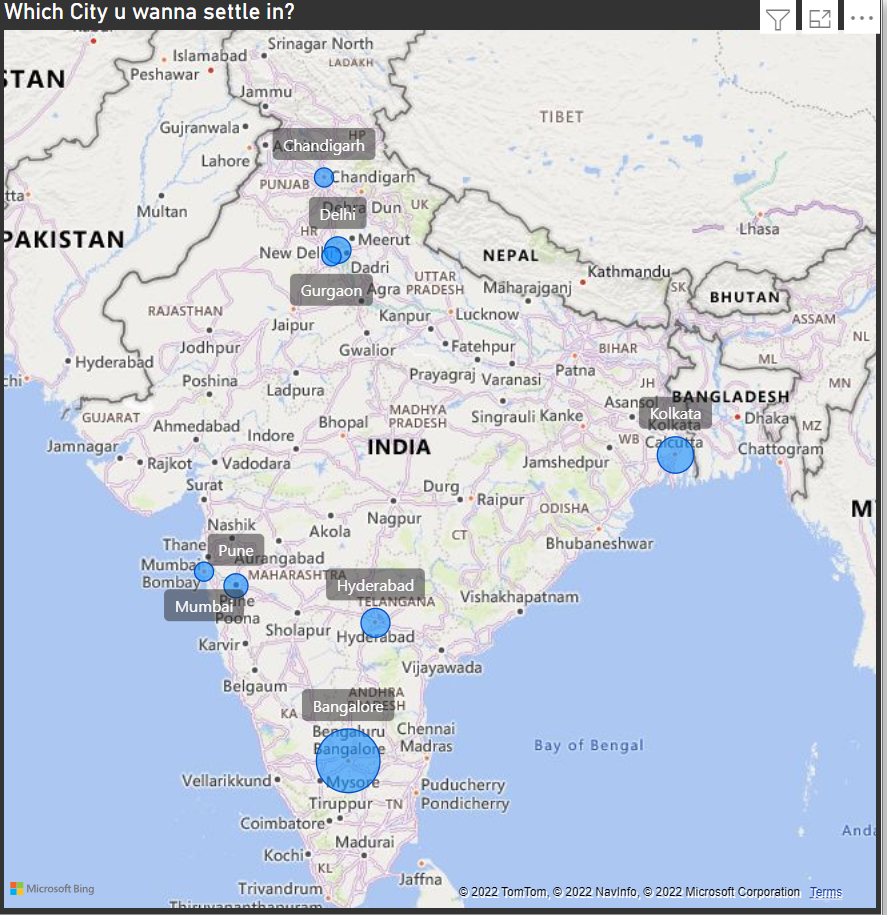
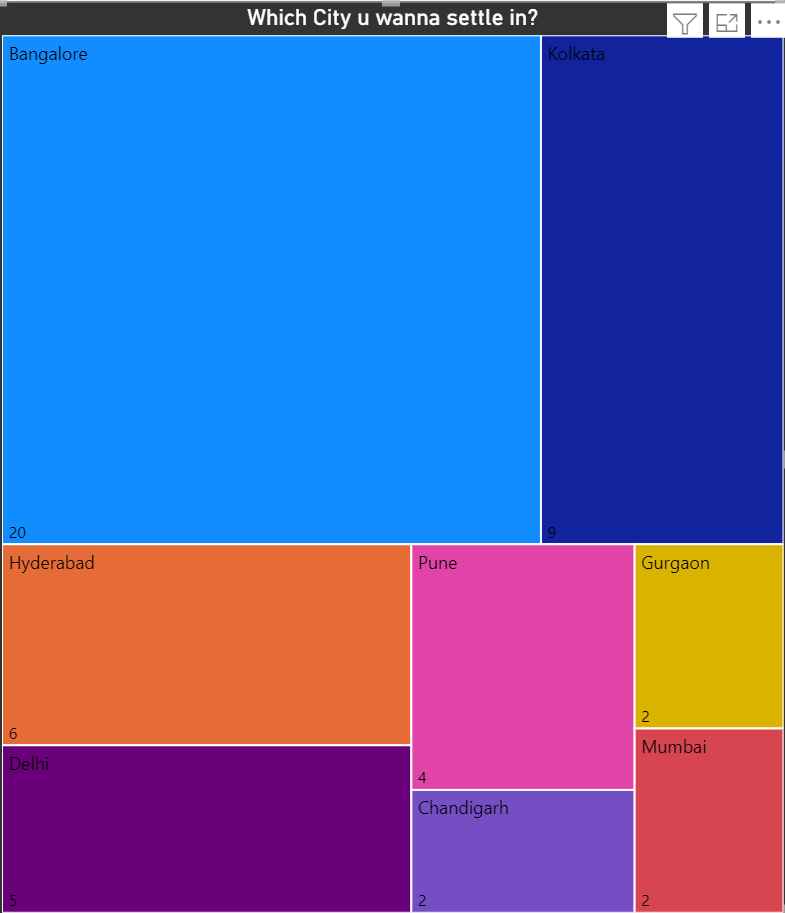
From the above graph its clearly visible that BTech individuals on an average desire a %change in salary in largest amount probably because they are more involved in technical sector and see their prospects in corporate world generally whereas one reason behind BSc grads expecting a lesser %change in salary can be due to the fact that BSc grads mostly delve into teaching and research sector,hence their expectations are lower in terms of salary.



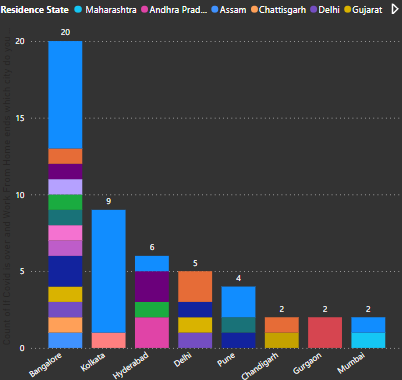
We have tried to figure out what average salary people are expecting on the basis of designations they want to get recruited as. So we can see that people who see themselves as Analytics Manager expect the highest salary of 15 LPA followed by Database administrators, ML Engineer, Data Engineer, Data Scientist, and Business Analyst whereas people who aspire to be Data Analyst expect the least salary on an average.



In general students are expecting a three times leap in their salary after switching to the field Data Science.

We asked the students which city do they want to get their job locations at if work from home ends. We got a quite interesting insight which showed most of them want to have Bangalore as their offline job location (nearly 40% of them want Bangalore) followed by Kolkata, Hyderabad, and Delhi which account for 18%,12%,10% of student responses respectively. This huge inclination towards Bangalore as a job location probably is due to the city’s huge job network as most of the companies are located there and consequently growth opportunities will be more there.



We drilled down the above tree map to find out which people from which residence states are trying to relocate to which cities after job and found out an interesting trend that people trying to work in a state are mostly living in and around in as expected but this trend goes for a toss for Bangalore where people from all sections of India are trying to relocate showing the true multicultural diversity of Bangalore which is a Metro city now in the truest sense.

**Conclusion**

From the above analysis (and in tandem with some outside research) we drew the following conclusion:

* People are transitioning into the Data Science Domain owing to massive amount of opportunities it provides be it job roles, responsibilities, challenges Career Growth Aspects Career Flexibility, and Remuneration. People from a variety of domains can make contributions to this sector owing to the fact that Data Science is an all-encompassing field with applications in IT, Healthcare, Defense, Manufacturing, and AI. So domain knowledge also can play a very important role here.
* Job-related Flexibilities-We know how Covid changed the workplace forever like with the Introduction of WFH. Data Science Jobs provide massive flexibility related to roles, job locations, Timings, and of course hybrid or work from home. These flexibilities allow people to spend more time to spend on themselves and their families. So win-win.
* People from many traditional sectors like Manufacturing, Construction, Chemical are also making a switch to this sector because of higher and more diverse opportunities coupled with higher remuneration than offered in those sectors.
* Extremely Innovative Work field with a massive amount of growth. New Inventions and Techniques in Data Science are forever changing the world for the better. For example, a massive increase in accuracy in detecting cancer, Weather prediction, Artificial Intelligence, Market Research, Customer Acquisition, etc to name a few. Young people love challenges and want to be part of this technological revolution that is Data Science.