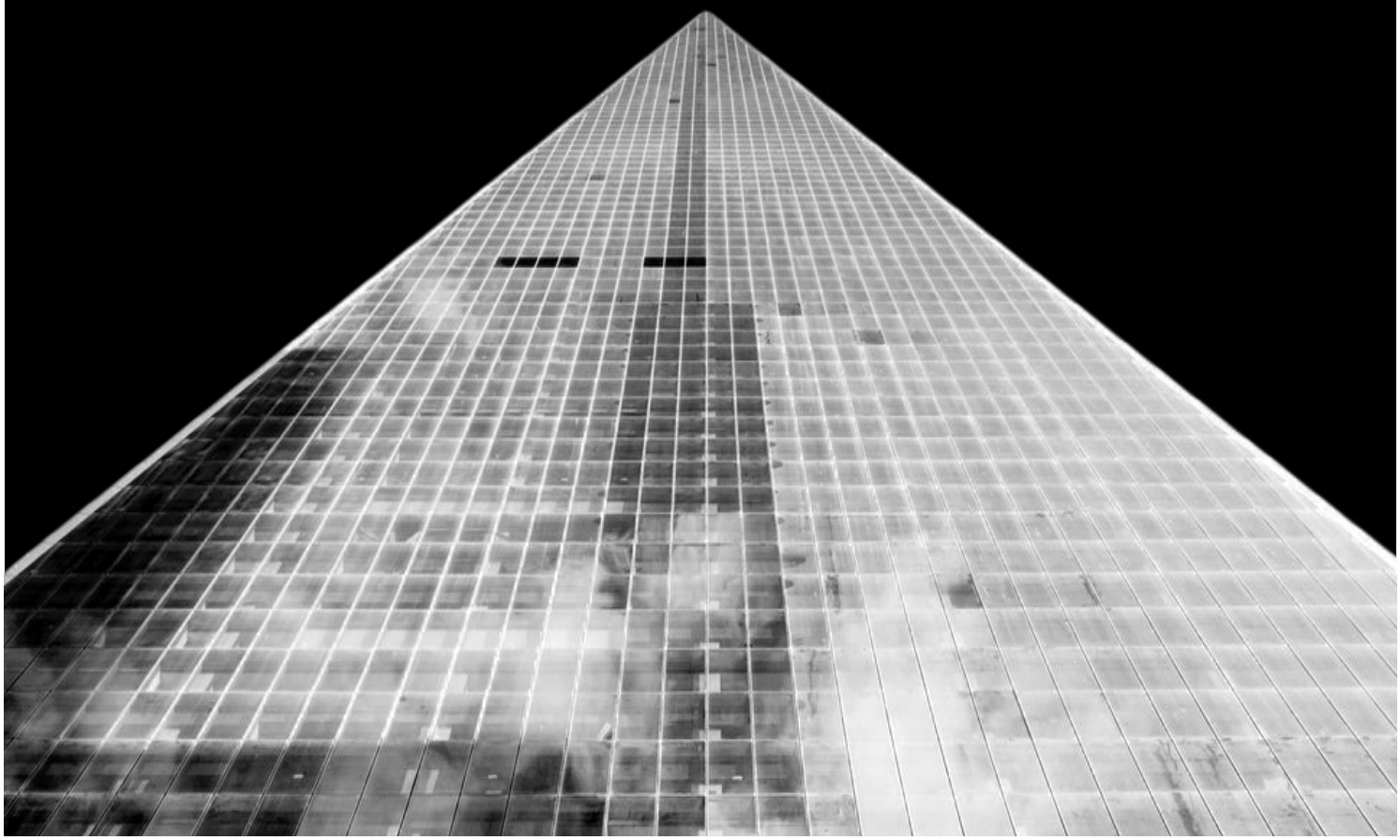
APRIL 2018 ISSUE 05

UpGrad

DATA STREAK

A monthly digest on all things Data



THE BIG SWITCH

JOURNEY TO THE HEAD OF ANALYTICS

AMIT GARG

Head of Analytics

CustomersFirst Now

I clearly remember the date even today, May 22nd, 2013. It was 5 am in the morning and I was all ready for my Skype discussion with my professor after my profile got shortlisted for a program at Indian Institute of Management, Bangalore.

I was nervous! I had gone through the program brochure and I had never heard or used more than 80% of the techniques. My profile was more of a Systems Analyst back then and my day to day responsibilities included business and systems analysis, requirement elicitation and writing requirement documents. At that time, I was based at the client's location in Connecticut, US and managing the development of a personal insurance policy administration system. The work was engaging..... But something was missing!

Right at that moment, I received a Skype call from the professor and the discussion began. There were talks around my current role, skill sets, past experience and why I was keen on joining such a course. There were a few technical and behavioural questions too.

To me, the discussion concluded reasonably well and I was satisfied that I had at least crossed one milestone. I was using data analysis to a limited extent and in odds and ends, however the intent was really to understand the entire spectrum and curiosity to unearth what all is possible using Analytics. Few days later I got a confirmation that I have been selected for the program. My joy knew no bounds.

There are a few instances that work as game changers in everyone's personal and professional lives. This was my turning point.

But this was just the beginning of entering an unknown territory, making some tough choices and most importantly, sticking to the choices that I have made.

I had to wind up everything from the US and move back to India within a month. A lot of support was required from my seniors within the company which I fortunately received. Soon after moving back to India, I joined office and the Data Science Program both at almost the same time, and very soon it felt like parking two cars in one parking space. This may sound like a funny analogy but I'm sure you got the idea! My personal, professional and academic life balance had gone for a toss. To be honest, there was no word like 'balance'. Everyday new learning. Statistical concepts, analytical algorithms, tools and languages and that's not it... Assignments at the end of each module. Did I forget to mention that I was an employee too for a high-demanding organization?

It looked like a big bubble that could burst anytime. But as I said, there were some tough choices made. The good news was that this phase was for a finite time and this thought kept my motivation levels high.

One after another, modules were finishing up and time was fleeting. I didn't even realize that I was able to cope with high stress levels and really upskilled myself in the analytics discipline that I had initially envisioned. I got more confident in my work and was able to add more value to the business. Not only that, I started seeing things differently. The trait of challenging the status-quo became even more dominating. Questions like, 'why is this happening', 'what's the real problem' etc. started bothering me more and the obvious inclination to search for answers to all the 'whys' was assumed to be hidden in data. **Data was the new God.**

At that moment, I was ready for my next BIG switch. More than anything else, it was a mental game. How much you are confident about yourself decides how much somebody else will be confident in you. As they say, 'there is no shortcut to success'; the more prepared you are, the more confident you will be.

Steve Jobs says – *"Life is too short to live somebody else's dream"*. So true!
Dream for yourself and give your everything to make that true. I wish all the best to all aspiring minds at UpGrad. You have already come so far by making some tough choices. Now, go and get It!

Work
Your Brain

1 You're about to get on a plane to Seattle. You want to know if it's raining. You call 3 random friends who live there and ask each if it's raining. Each friend has a 2/3 chance of telling you the truth and a 1/3 chance of messing with you by lying. All 3 friends tell you that "Yes" it is raining. What is the probability that it's actually raining in Seattle?

TIPS & TRICKS TO WORK AROUND WITH REGRESSION

SUMIT KUMAR

Content Strategist

UpGrad

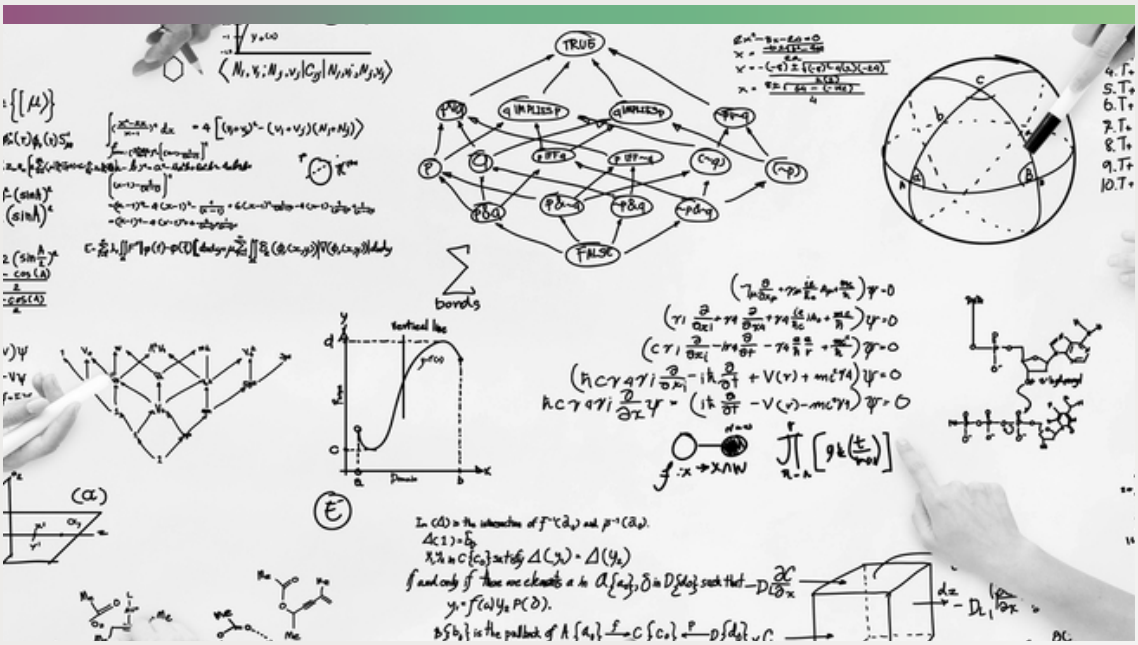
Linear and Logistic regression are two of the most used algorithms and they are also widely prefered by analysts in the industry too.

The main reason for being so popular is because of its simplicity. Linear and Logistic regression are some of the few algorithms which most of the students and professionals learn to enter into the world of predictive analytics. But do we have a proper knowledge of this most used algorithm?

Let's look at some tips to work with regressions and unleash its power of prediction and classification.

RESEARCH:

Most of the amature analysts commit this mistake of jumping directly into the predictions using the built in functionality, they often don't try to understand the problem they are trying to solve with the algorithm, now this is the very first mistake we all can commit. Regression is not all about just running the model and getting results. We can predict with the best



accuracy using regression only when we have properly did our research before starting regression analysis.

a. First, always understand the problem your are trying to solve.
b. Always perform EDA to understand the data you are going to deal with.
c. Spend most of your time in understanding the importance of the variables in the given data.
d. Having proper domain knowledge can help you identify the most important variables.

DATA PREPARATION:
Now this is the most important step while going for regression and it is used to take almost half of your time in the whole process of model building.

a. Data cleaning: Missing values and outliers are the most common treatment we need to test and treat; keep in mind missing values and outliers are like default components which come with any standard dataset, so always check them even if not instructed by your boss or head.
b. Check if missing values are needed to be imputed or can we afford to drop the missing rows. In missing values treatment, imputing missing values is not always a good solution sometimes dropping them is good since imputing them with mean or median will create a little bias in the data.
c. Derived variables: Always check if we really need derived variables, always keep in mind introducing more number of variables in the model will not make your predictions more accurate, so always discuss with your manager that if adding a derived variables will actually help in improving the prediction or not.
d. Variable reduction: As I previously mentioned increasing the number of variables will not increase the accuracy so it's good to reduce the variables and retain only the best. Some simple variable reduction techniques we can use:

- Drop the variable with large missing values (40%-50%)
- Low Variance - Continuous variable with low variance should be removed.
- High Correlation - Variables with high correlation with other variables other than dependent variables can be dropped.
- Dropping variables doesn't make any business sense.

MODEL BUILDING:
Now once our data is ready we can start building a predictive model but always remember some tips:

a. First run the model with all variables.
b. Then check which all variables are insignificant.
c. Then perform simple regression between dependent variable and the insignificant variables one by one and check the coefficients.
d. Finally remove only those variables that are having very less coefficient values even using simple regression.

CHECK FOR ASSUMPTIONS:
Now it's time to check for assumptions while running linear regression.

a. Your data cannot have any major outliers, or data points that exhibit excessive influence on the rest of the dataset.
b. Variable relationships exhibit (1) linearity – your response variable has a linear relationship with each of the predictor variables.
c. Your data shows an independence of observations, or in other words, there is no autocorrelation between variables.
d. Your data demonstrates an absence of multicollinearity.
e. Your data is homoscedastic.
f. Your residuals must be normally distributed.

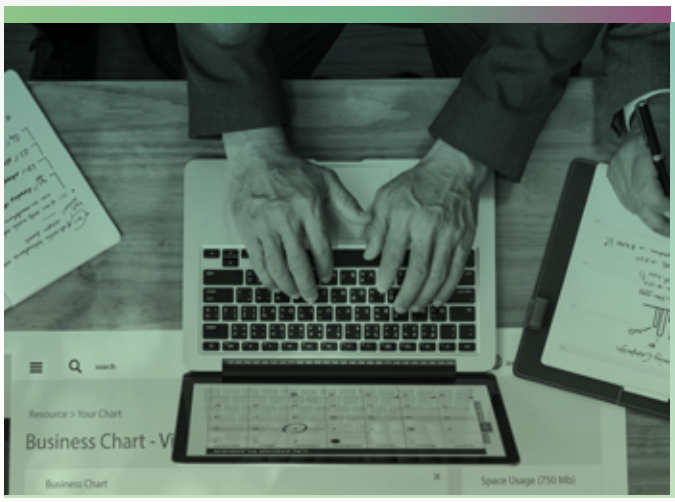
VALIDATION OF THE MODEL:
Now we need to check if the model is working fine with the data that was unseen to it. For that:

a. Test your model against test data and see for r squared and adjusted r squared values.
b. Check the result, if it is actually making sense or not.
c. Use cross validation for better results.

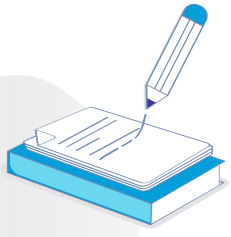
COMMON PROBLEMS & ITS SOLUTIONS:
If we will not take care of all the assumptions in the regression we may often get wrong results, so it's better to always check:

a. If your data is heteroscedastic, you can try transforming your response variable.
b. If your residuals are non-normal, you can either (1) check to see if your data could be broken into subsets that share more similar statistical distributions, and upon which you could build separate models OR (2) check to see if the problem is related to a few large outliers. If so, and if these are caused by a simple error or some sort of explainable, non-repeating event, then you may be able to remove these outliers to correct for the non-normality in residuals.
c. If you are seeing correlation between your predictor variables, try taking one of them out.
d. If your model is generating error due to the presence of missing values, try treating the missing values, or use dummy variables to cover for them.

Regression looks simple and easy but to get the best results using it is really a hard task. The only way to land on the best results is by always following the steps and carefully examining them at each step.



Work Your Brain



2 By using number 7,3,7,3 can you get number 24 by using any mathematical signs (+, – , x, /)?

CRACKING THE INTERVIEW

By UpGrad

One of the most common questions asked in every interview these days, is “Describe a project you have worked on.”

This might not be the most difficult question in an interview but it’s still a tricky one.

Here is what we suggest you can do to prepare yourself for such a scenario.

Identify any one project from your resume that you are confident about and have a thorough understanding of. Then, work on these simple pointers:

- *Understanding of the problem that is on-hand and your ability to explain the same*
- *Basic assumptions and outliers*
- *The model that you chose and logic behind choosing that model*
- *Result and accuracy achieved*
- *Impact that it had on business*

Here are a few reasons why interviewers ask the above question.

They want to see how well you can manage a project or a situation, what your approach to dealing with challenge is and how your skills would help you successfully lead a project. They also want to know what your work ethic is like and gain an insight into how you handle stress.

Coming back to the preparation bit.

First, you need to prepare an answer in advance. It’s very hard to give a well thought-out, five-star answer if you haven’t done prep before your interview. To get started with prepping a response to this interview question, write out a list of all the important projects you’ve worked on in yourcareer. Then, note what the goal was for each project and what part you played in bringing each project across the finish line. Then, note the outcomes of all the projects.

The next step is to hone in on the project (or projects) you’ll discuss in an interview situation. Use the STAR method (Situation, Task, Action, Result) to create a concise (yet thorough) answer. Make sure your answer demonstrates your abilities with setting priorities, making decisions, hitting deadlines and delegating tasks (if you’ve been in a management role)

Consider the points below while preparing for the answer:

Choose the right example/project -
Focus on something recent, successful and relevant to the applied position.

Be specific -
Talk (be through) about your process.

Explain your role clearly and talk tangible outcomes -
Go ahead and quantify your answer as much as possible! Also share your learnings.

Following these tips on how to explain your projects will help you articulate and present your work in a much structured manner and will give the interviewer a clear understanding of the work you have done and the models you have worked with.

HireSmart

RECRUITMENT DRIVE

In a series of many firsts, UpGrad conducted its first offline placement drive on 15th March, 2018 for the learners of Data Science cohort 2. With over 7 companies participating, UpGrad Xchange at Bengaluru saw some of the biggest names in the Analytics industry like Zivame, Fractal Analytics, Kantar Analytics and Animaker to name a few.

Amongst an application pool of 150, around 90 learners got shortlisted and 7 of them are currently holding offer letters in their hands. More offers are expected to roll in soon. It was great to witness participation from both, the learner's and the company's end and we expect to see more such satisfied faces in our future placement drives.

OUR RECRUITMENT PARTNERS

Fractal | Niki.ai | GAIN THEORY | zivame | Animaker

GLOBAL analytics | KUMARAN SYSTEMS | springML | FUNDSINDIA | nabler

STUDENT SUCCESS STORY

Q1. What was your primary motivation to learn Data Science?

The motivation came from multiple sources. A tough placement season, seeing how my firm leverages data and this quote from those fancy B-school admission videos pushed me to take this up. “When you know what is it that you don’t, that moment it hits you what you need to do.”

I quickly picked up my mark sheet, saw my abysmal statistics score (embarrassing after calling yourself an Economics Graduate) I thought I must take the plunge and learn data analytics.

MOULIK SHRIVASTAVA

Business Analyst

Bain & Company

Q2. How did you come to know about UpGrad?

The Google Ads embedded onto most websites I visited after a quick Google search (obviously based on the trove of data Google has on me) were effective enough for me to explore the program.

Q3. Why did you choose UpGrad to learn Data Science?

The industry oriented case studies, the exceptional learning platform and backing of a known institute. Convincing parents to pay a mammoth sum for an online course is difficult, but the fact that it was a PG Diploma added a lot of weight.

Q4. What impact did the program have on your career transition?

In a few words : 2X CTC Growth !

Q5. What extra steps outside the program learning did you take to escalate this career transition?

Since the beginning, I knew the difference between career services & placement. Also, after a quick glance on the Google sheet showcasing the cohort's other learners, I knew that a shot at the best roles would be far fetched, especially when there are so many hardcore coders in the cohort and my non-tech background.

Thus, I did the following :
Actively looked up on LinkedIn & Glassdoor, putting in keywords relevant to data analytics. Did mini projects at my workplace to build credibility (asked for LinkedIn recommendations) on my data skills. These projects greatly helped connect business and data objectives.

Q6. How has your experience been so far with the program? Would you recommend the program to any of your friends/colleagues?

Absolutely fantastic! While deadlines and vast content gets hectic, time spent on the learning platform, whatsapp with learners, and case studies is time well spent and knowledge picked up. The mentorship through industry veterans, BaseCamps (with Kindle like prizes) and student mentors make the course extremely engaging.

As far as recommending goes, I would definitely endorse the programme for its rich content and comprehensive approach to Data Science.

Work Your Brain

3 Mention which lamp is the brightest of all?
Lamp A is dimmer than Lamp B
Lamp B is brighter than Lamp C
Lamp C is as bright as Lamp D
Lamp B is brighter than Lamp D
Lamp D is brighter than Lamp A

4 You have a 100 coins laying flat on a table, each with a head side and a tail side. 10 of them are heads up, 90 are tails up. You can't feel, see or in any other way find out which 10 are heads up. Your goal: split the coins into two piles so there are the same number of heads-up coins in each pile.

CAREER TRANSITIONS

Hari Krishna Reddy

Career Break

Consultant

Fractal

Suman Ashwin

Product Management

Quantum Leap Laboratories

Technical Project Manager

CCP

Sushil Malani

Career Break

AVP Customer Engagement

Cigna TTK Health Insurance

Ankur Srivastava

Senior Test Lead

EmQos

Data Analytics Intern

MAATRITVA Enhancing Motherhood

Bhagavathi C

Intern

Nikhil Analytics
Performance Driven By Values

Data Analyst

GAIN THEORY

Renu Mishra

Software Developer

Capgemini

Data Governance Developer

@WalmartLabs

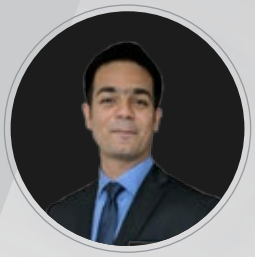
Amaresh Dhal

Consultant

J.P.Morgan

Data Scientist & AI Engineer

IBM



Moulik Shrivastava

Business Systems Analyst

AMERICAN EXPRESS

Business Analyst

BAIN & COMPANY

SOLUTIONS

Every organization wants its' employees to possess strong analytical skills which refers to the ability to gather data, break down a problem, weigh pros and cons and reach logical decisions. Employees who have these skills help companies overcome challenges, or spot issues before they become problems. Employers like to assess your analytical abilities during interviews.

Hope you find these these sample analytical skills interview questions helpful which were asked to our learners during their interview.

1 You only need 1 of your friends to be telling the truth for it to be raining in Seattle. It's fastest just to calculate the odds that all 3 are lying, and it's not raining. Each friend has a 1/3 chance of lying. Multiply the odds together... you get 1/27 (1/3 * 1/3 * 1/3). We're not done yet though... 1/27 is the probability that all 3 friends lied at the same time. The probability that at least 1 told you the truth? 26/27 or around a 96% that it's raining in Seattle.

2 7x ((3/7) +3) = 24

3 Lamp B

4 You want equal number of heads in each pile. There are currently 10 of them. You don't know which but it doesn't matter. All you have to do... take any 10 coins out of the 100, put them into a separate pile, and flip those 10 over.
That's pile #1.
Pile #2 is the remaining 90 coins, unflipped. Just leave them.
You're done.

CLICK HERE TO SEND US YOUR FEEDBACK

FIND US HERE:

To share your stories/articles/blogs,
write to us at pgdds@upgrad.com

UpGrad Education Private Limited, Nishuvi, 75, Dr. Annie Besant Road, Worli, Mumbai – 400018.