# Initial Posts

Does anyone have any experience with Data Exploration & Analysis? If so, what type of data have you worked with and what were some challenges you faced with that data?

For me, I'm currently a BI Developer for a Pediatric Hospital in Texas. I've been in the Healthcare industry for over 10 years but got into Data & Analytics in 2017. Healthcare data in my opinion is very messy. Multiple departments use the system differently, workflows are always changing, NULL values are everywhere, data sets are very large and hard to work with, and business logic and definitions are debated constantly. Although challenging, I love working with Healthcare data and knowing that the information created can potentially help an organization that's main goal is helping sick children.

**Pandas**

Found a really good series talking about Pandas in regard to data manipulation and cleaning via the Pandas package in Python. Coming from a SQL background, this package seems very intuitive to use to clean data.

[Python Pandas Tutorial (Part 1): Getting Started with Data Analysis - Installation and Loading Data - YouTube](https://www.youtube.com/watch?v=ZyhVh-qRZPA&list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS)

**Project Topic**

Anyone started on their project yet or have an idea on a topic for conducting Exploratory Data Analysis on?

For this course I've decided that I would pick a topic that pertains to something that both me and my wife care a lot about which is Autism Spectrum Disorder (ASD). I'll be trying to find variables which can potentially predict ASD rates per state in the U.S.

Before our first child was born earlier this year, my wife was a behavior therapist at Child Study Center where she would provide early intervention behavior services for children with ASD in hopes that they would be independent once they were older. We also were involved in a program to give parents with ASD children a much needed break 1 weekend every month. Having a child with ASD is a full-time 24/7 job with no breaks which is very stressful so this government program aimed at giving parents a break by allowing the parents to drop their kids off with licensed behavior therapists.

**Public Health Data**

If anyone is looking for a great source of public health data, the CDC has a tool to pull free open data sets online called wonder (link below). If your term project has anything to do with public health this looks like it would be a very valuable source. Hope it helps!

**Exploratory Data Analysis (EDA)**

Exploratory Data Analysis is an important step when creating a predictive model. It is the step where the Data Scientist or analyst analyzes the data to see if there are any patterns or correlations in the data. This can be done by running statistical analysis and also by creating visualization of the data. Knowing patterns and trends in the data will help with feature selection and model selection.

# Replies

Nice to meet you Amelia! Yes, I was thinking the same thing and am looking forward to learning more about Data Exploration & Analysis in regards to Data Science.

Hi Deborah, nice to meet you. What sort of industry have you worked in previously to report on KPI's?

*Regarding the question, “What is Exploratory Data Analysis?”*.

Exploratory Data Analysis (EDA) in regards to Data Science is the step in which the data is investigated and understood. Some questions hopefully answered during this step in the process might be:

* How does the data get stored in the database?
* What workflows cause the observed data to be as it is?
* Are there any data anomalies? If so, why/how?
* Are there any patterns in the data?
* Do certain variables seem to correlate with an outcome?
* What does the data look like visually?
* Are variables statistically significant?
* Etc.

*Regarding Data Science Methodologies:*

Hi Amelia, I too prefer the CRISP-DM methodology and applaud its consideration of iteration which, in my experience, typically is what occurs in the majority of the analytics projects that I've worked on. For me personally however, once there is a problem established which requires data analysis, I usually tend to start at the "Data Understanding" step where I query the database to form questions about the data and then move backwards to the "Business Understanding" step where I reach out to operational or workflow experts who can hopefully help me answer questions about why the data is the way it is. Not sure if this breaches the CRISP-DM methodology rules but other than that it looks like this is the methodology I use in my day-to-day.

*Regarding question about Teams Word document instructions:*

I asked about this as well. Per Dr. Parajulee, you would just copy your 5 initial posts and 5 replies and paste them into a word document. Either on Sunday (last day of week) or Monday (first day of week) in the teams channel, start a new conversation and click the 'attach' icon and attach the word document (see below). Then, tag Dr. Parajulee using the @ tag.