EDA

- Exploratory Data Analysis

Mass Street University

PER EDUCATIONEM PROGRESSUS

EDA

Traffic was horrible back then.

Mass Street University
PER EDUCATIONEM PROGRESSUS

EDA

- I walked into my DS program with the following
  - One year of stats
  - Math training up to Diff Eq
  - Completed Andrew Ng's and numerous other math dense MOOCs on machine learning
  - Had completely read several other math dense books on machine learning
- YOU DON'T NEED ANY OF THIS ANYMORE!

Mass Street University

PER EDUCATIONEM PROGRESSUS

EDA

- Today's data scientist need to understand their tools, but not implementation details.
- Algos need proper care and feeding.
- Algos suffer from the same problem all algos do. GIGO.
  - Garbage in. Garbage out.
- Modern algos correct for things we had to manually deal with just 5 years ago.

Mass Street University

PER EDUCATIONEM PROGRESSUS

Oh lord. Here we go again.

Mass Street University

PER EDUCATIONEM PROGRESSUS

EDA

- We use EDA to understand the data that we're going to analyze.
- This is where you earn your junior P.I. merit badge.

Mass Street University

PER EDUCATIONEM PROGRESSUS

EDA

THE FIRST STEP IN EDA IS TO PERFORM A RECORD COUNT!!

Mass Street University

PER EDUCATIONEM PROGRESSUS

Here. You’re gonna be a while.

Mass Street University
PER EDUCATIONEM PROGRESSUS

These stories may seem like they don’t add value.

I tell them because it’s WILD out in these streets and you need to know what you’re getting into.

Mass Street University

PER EDUCATIONEM PROGRESSUS

Dirt. He means dirt.

Mass Street University

PER EDUCATIONEM PROGRESSUS

These stories may seem like they don’t add value.

I tell them because it’s WILD out in these streets and you need to know what you’re getting into.

Mass Street University

PER EDUCATIONEM PROGRESSUS

These stories may seem like they don’t add value.

I tell them because it’s WILD out in these streets and you need to know what you’re getting into.

Mass Street University

PER EDUCATIONEM PROGRESSUS

EDA

- When you’re crunching numbers, you need to understand how much horsepower is at your disposal.
- Be careful. Some algos explode the dataset and will choke your machine if you’re running an analysis on a single box.

Mass Street University

PER EDUCATIONEM PROGRESSUS

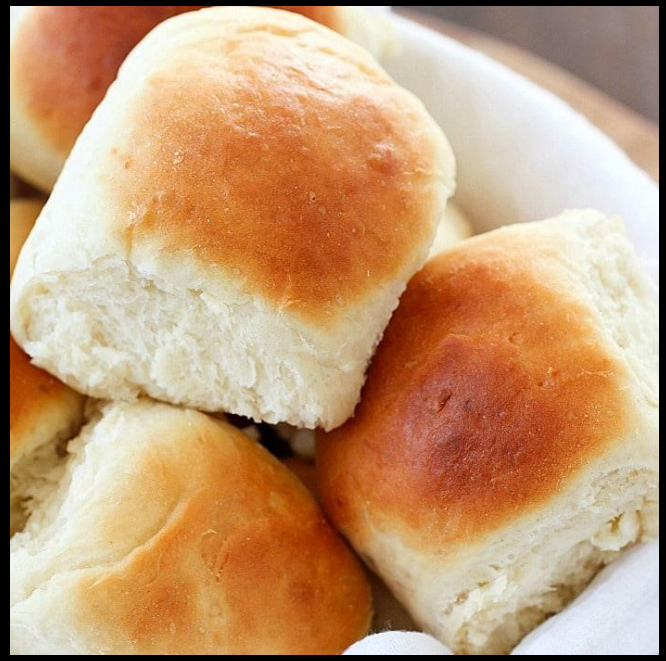
EDA

- The next step in the EDA process….

Mass Street University

PER EDUCATIONEM PROGRESSUS

EDA



This is how I roll.

EDA

- Do a data summary
- Summaries contain basic statistical information
- You'll need to do some viz because some stats are hard to understand as just numbers.

Mass Street University

PER EDUCATIONEM PROGRESSUS