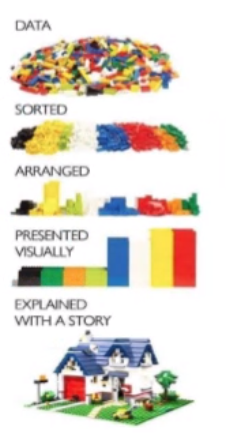
**Week 2 Notes**

**Day 2**

**Intro to Data Analytics**

Data Analytics and Python skills are a highly sought after job by employers

**GET INFO OFF THE VIDEO FOR** THE definition

**Four main categories** – *descriptive* *analysis* – identify trends and relationships

*Diagnostic* – determine why something happened

*Predictive* – makes predictions about future outcomes

*prescriptive* – determine the optimal course of action

*Five V’s* – volume, value, variety, velocity, veracity

*Data analytics is imperative in*: transportation, logistics, delivery, supply chain management, insurance and assessment, fraud and risk detection, security and safety industry

*Who are Data analysts* – *key traits* – perceptive, analytical, and detail oriented—work varies depending on the data you’re working with –

Skillset – programming, python, data wrangling – excel, APIs, database management, data visualization

Python – general purpose highly versatile, software development, data science, scripting, staple in data science because statistical calculations, data vis, machine learning algorithms, manipulate and analyze data

Additional Skills – version control (git), different database technology, statistical processes, basic understanding of front end tech, project management (agile), excellent communication skills

Diagram

Description automatically generatedChart

Description automatically generated with low confidence

CRISP-DM – cross industry standard process for data mining – <https://www.datascience-pm.com/crisp-dm-2/>

Data Wrangling – cleaning and unifying messy and complex data sets for easy access and analysis EX: joining data sets together, finding gaps, cleaning data, identifying outliers) Goals – enrichment, actionable data, time saving, focus on analysis, decision making

**Diagram

Description automatically generatedStep one** – Discovery – find data that addresses your question

**Step 2** – Structuring – take raw data and transform it to what you can work with

**Step 3** – Cleaning – remove outliers that can skew results when analyzing data

**Step 4** – Enriching – decide if you need to add by combining raw data with additional other sources if original source is not adequate

**Step 5** – Validating – make sure the data is high quality

**Step 6** – Publishing – Creating analysis

Data Wrangling – modifying and/or preparing data

Data cleaning – removing unnecessary data

**Day 2**

Table

Description automatically generatedExcel and Data Analytics

*Data Wrangling* – processing and organizing data

*Excel* – program that’s used for recording, analyzing and visualizing data in the form of the spreadsheet

Graphical user interface, text, application

Description automatically generated*Benefits* – performs various math functions on large data sets, you can search sort and filter, beautifies data and present with charts and tables

**Day 3**

Excel Functions – Math, logical, statistical

https://support.microsoft.com/en-us/office/excel-functions-by-category-5f91f4e9-7b42-46d2-9bd1-63f26a86c0eb

Math Functions – Sum adds cells (A2:A10) would add a2 and a10; SUMIF – calculates the sum in a range based on true/false conditions; SUMIFS – calculate the sum of multiple ranges based on one or more t/f condition;

Logical – if – make logical comparisons between a value and what you expect (true or false);

If (equal to), IF (greater than), AND determine if all conditions are true mus be true or false; and using if; OR only one conditions must be true; OR WITH IF lets you check multiple conditions for IF; IFS – returns a value based on the first TRUE condition, XOR returns a logical

Statistical Functions – AVERAGEIF – Combine averages from different cells, AVERAGEIFS calculates the avg of a range based on one or more t/f conditions, COUNT calculates the number of cells used within a range that have numbers, COUNTA – counts the number of cells used within a range

COUNTBLANK – calculates the number of cells used within a range that are blank, COUNTIF – calculates number of cells as specified have to use a $ for absolute values to be counted, COUNTIFS – counts cells in a range based on one re more t/f conditions, MEDIAN MODE, MODE.SNGL, STDEV.p – calculates the standard deviation for the entire population, STDEV.s – calculates the Standard Deviation for a sample, VLOOKUP – allows searched across columns