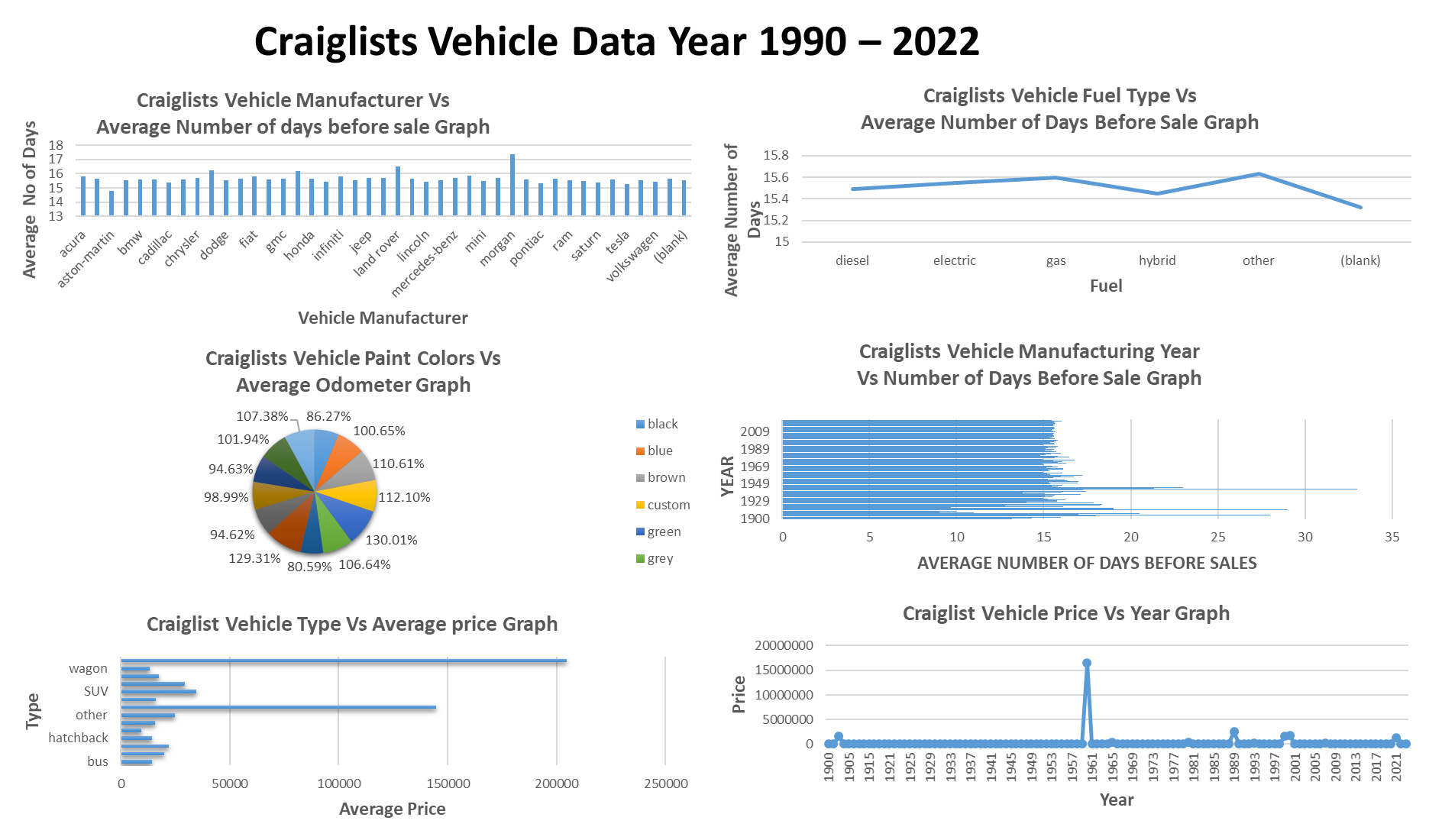
**DSEAFRICA WK 5 PROJECT:**

**WEEK 5 ARTICLE**

**DATA ENGINEERNG FOR BEGINEERS- A STEP BY STEP GUIDE** Data Engineering is the data management discipline that deals with designing, developing and maintain systems to handle data in large scale by collecting, integrating raw data, translating, and validating data for analysis and other applications. It is the foundation for the application of data science in reality, as it delivers formatted, scalable secure data, programming and system architectural skills.

Data Engineering involves the creation of data pipelines that focuses comprehensively on data collection, processing, storage from various sources using Extraction, transformation and loading (ETL) design to build data warehouse, data modelling, relational and non- relational database and query execution to ensure conversion of data into usable information for interpretation, ensure unified accessibility of data and to enhance driven decisions.

Data Engineering Sources

1. Structured data: it is a well-defined data in rows and columns (schema) example relational database and spreadsheet.
2. Semi- structure data: partially structured but without schema (rows and columns) and provides additional Information example e-mail, Zip files.
3. Unstructured data; data that lacks a well-defined schema example include images, videos, website data.

Data Engineering Skills

1. Knowledge of programming g languages such as SQL, Python, Machine learning, data processing, ETL Techniques, Java, Scala.
2. Data structures and Algorithm
3. Database management system
4. Data Exploration (PySpark, Pandas, NumPY, Spark, one cloud, big querry, snowflake)
5. Data warehousing
6. Data visualization tools

**Why Data Engineering?**

As technology advancement increase it created the demand for talented professionals with passion and problem-solving approach to handling data by the following reason:

1. High demand and marketability
2. Diverse opportunities
3. Continuous learning and growth
4. Impact and value

**STEP BY STEP GUIDE TO DATA ENGINEERING**

STEP 1: Understand the basics in data engineering

Research and understand the responsibilities, expectations and market demand for data Engineers in various industries.

STEP 2: Develop data engineering skills and tools such as:

Strong analytical and problem-solving skills, familiarity with SQL, proficiency in programming language (SQL, Python, Java, programming, ETL, Pycharm, Jupyterlab, Spyder

etc.), experience in data processing tools (Hadoop, Spark) and ability to effectively communicate complex technical concepts.

STEP 3: Seek avenues to explore the new skills through:

Open-source contributions to learn from experienced professionals and establish a presence professionally example kaggle, twitter, explore personal projects to apply and showcase skills by internships positions.

Step 4: Display skills on social media platforms by connecting with like minds to build on line presence by posting on projects on websites, twitter, LinkedIn, GitHub.

STEP 5:

Practice interview questions

STEP 6:

Get real world experience to apply knowledge and skills through internships, work opportunities to solve challenges.

Step 7: Stay updated with the latest advancements in data engineering technologies and continuous learning and upskilling, follow industry trends and advancements, participate in online courses, webinars and conferences to stay updated, networked and upskilled to maintain relevance.