



**Department of Computer Science and Engineering (Data Science)**

**Subject: Applied Data Science (DJ19DSL703)**

**Experiment -4**

**(Data Preparation Best Practice)**

**Name: Jhanvi Parekh**

**SAP ID: 60009210033**

**Batch: D11**

**Aim:** To complete Data Preparation Best Practice.

**Theory:**

Data preparation is a crucial step in any data science project, and it involves various activities, including gathering suitable data, determining key performance indicators (KPIs), and creating dashboards for business stakeholders.

Writing Key Performance Indicators (KPIs) for a data science project involves defining measurable metrics that align with the project's objectives and help assess its success.

Proof of Concept (POC) refers to a demonstration or preliminary test conducted to assess the feasibility and viability of an idea, concept, or solution. It serves as an initial evaluation to determine if the idea is both possible and worth pursuing.

**1. Gathering Suitable Data for a Data Science Problem:**

**a. Understand the Problem:**

- Clearly define the problem you are trying to solve with data science. Understand the business objectives and goals.

**b. Identify Data Sources:**

- Determine where you can obtain the necessary data. Sources may include databases, APIs, external data providers, and internal data repositories.

**c. Data Relevance and Quality:**

- Ensure that the data you gather is relevant to the problem at hand. Irrelevant data can lead to noise and inefficiencies.
- Assess the quality of the data by checking for completeness, accuracy, consistency, and any missing or duplicate values.

**d. Data Collection Plan:**

- Develop a plan that outlines how you will collect data. This includes specifying the data sources, collection methods, frequency, and responsible individuals or teams.

**e. Data Licensing and Compliance:**



### **Department of Computer Science and Engineering (Data Science)**

- Be aware of data licensing and legal considerations. Ensure that you have the rights to use the data for your intended purpose.
- Comply with data protection regulations (e.g., GDPR) if applicable.

#### **f. Data Sampling:**

- For large datasets, consider taking a representative sample for initial exploration and analysis. This can save time and resources.

## **2. Determine All Key Performance Indicators (KPIs):**

#### **a. Understand Business Objectives:**

- Collaborate closely with business stakeholders to understand their goals and objectives. KPIs should align with these business goals.

#### **b. Identify Relevant Metrics:**

- Determine which metrics and measurements are relevant to assess the success of your project.
- Prioritize a small set of critical KPIs to focus your efforts.

#### **c. Define Measurement Methods:**

- Clearly define how each KPI will be measured. This includes specifying data sources, calculation methods, and any necessary formulas.

#### **d. Set Baselines and Targets:**

- Establish baseline values for KPIs to provide context. Determine what constitutes success by setting target values or thresholds.

#### **e. Monitor and Report:**

- Implement a system to monitor KPIs continuously. Automate data collection and reporting when possible.
- Create dashboards or reports to visualize KPI trends and share them with stakeholders.

#### **f. Iterate and Refine:**

- Be prepared to iterate on your KPIs as your project evolves. If certain KPIs are not providing actionable insights, consider refining or adding new ones.

## **3. Business Stakeholders POC Dashboard:**

#### **a. Identify Stakeholder Needs:**

- Collaborate with business stakeholders to understand their requirements and the specific insights they need from the data.

#### **b. Dashboard Design:**



### Department of Computer Science and Engineering (Data Science)

- Design the dashboard with a focus on user experience and simplicity. Ensure that it is easy to navigate and understand.
- Choose appropriate visualization types (e.g., charts, graphs, tables) based on the nature of the data and the insights you want to convey.

#### c. Real-Time or Periodic Updates:

- Determine whether the dashboard should provide real-time or periodic (e.g., daily, weekly) updates.

#### d. Data Security and Access Control:

- Implement security measures to ensure that sensitive data is protected. Define access control rules based on user roles and permissions.

#### e. User Training and Support:

- Provide training and support to business stakeholders to help them make the most of the dashboard.

#### f. Feedback and Iteration:

- Continuously gather feedback from stakeholders to improve the dashboard over time. Ensure that it remains aligned with their evolving needs.

### Lab Assignment:

1. Write concise KPI statements that clearly describe what each metric measures and how it aligns with project objectives. Use specific, action-oriented language. Each KPI statement should follow the SMART criteria:
  - Specific: Clearly define what is being measured.
  - Measurable: Specify how the metric will be quantified.
  - Achievable: Ensure that targets are realistic.
  - Relevant: Ensure that the metric is directly related to project goals.
  - Time-bound: Define a timeframe for achieving the target or benchmark.

### Example

:

#### KPI: Customer Churn Rate

**Definition:** The percentage of customers who stopped purchasing from the company during a specific time period.

**Measurement:** (Number of customers lost in a period / Total number of customers at the beginning of the period) \* 100.

**Target:** Reduce churn rate by 15% within the next six months.

2. Create a POC dashboard using Power BI for the sample dataset of the chosen data science project.



## Department of Computer Science and Engineering (Data Science)



### Welcome to PhoneNow

#### Key Performance Indicators

1. Increase tech support capacity for Fiber Optic customers and lower tech tickets per customer to 0.5
2. Increase sale of 1 and 2 year contracts by 5% each
3. Yearly increase of automatic payments by 5%

#### Churn Dashboard



- Demographics
- Customer Account Information
- Services

#### Customer Risk Analysis



- Internet Service
- Type of contract
- Payment Method



### Churn Dashboard

1869

Customers at risk

2173

# of Tech Tickets

885

# of Admin Tickets



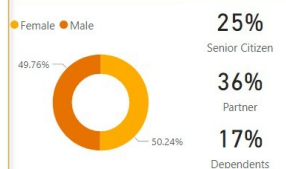
\$2.86M

Yearly Charges

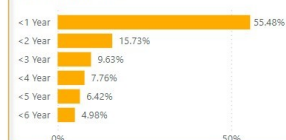
\$139.13K

Sum of MonthlyCharges

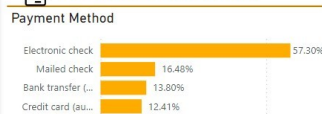
#### Demographics



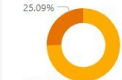
#### Subscription Time



#### Customer Account Information



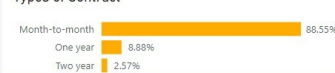
#### Paperless Billing



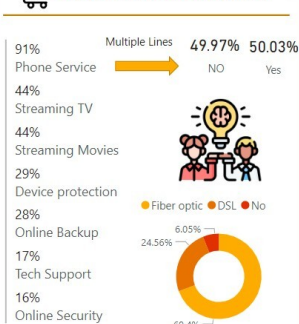
#### Average Charges

\$74.44 Monthly  
\$1,531.80 Total

#### Types of Contract



#### Services Customers Signed up for





**Department of Computer Science and Engineering (Data Science)**



## Customer Risk Analysis

### Risk of Churn

- ☐ No  
☐ Yes

### Internet Services

- ☐ DSL  
☐ Fiber optic  
☐ No

### Months Subscribed

0 72

### Contract Type

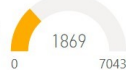
- ☐ Month-to-month  
☐ One year  
☐ Two year

7043

Total Customer

26.54%

Churn Rate %



\$16.06M

Yearly Charges

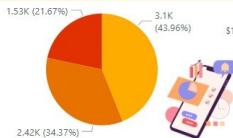
3632  
Admin Tickets  
2955  
Tech Tickets



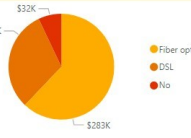
### Churn By Type of Internet Services



### #of customers by Internet services



### Sum of Monthly Charges



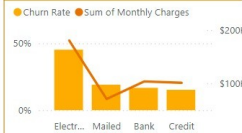
### Type of Contract



### Years of Contract



### Churn by Payment Method



Page 1

Page 2

Page 3

