SOLUTION PLAN

1) Preprocessing done

- **Inspected data (info, describe, null & duplicate checks)**

- **Removed missing values, irrelevant columns**

- **Datetime parsing → datetime_of_collision**

- **Converted key columns to categorical / string types (Type casting)**

- **Separated numerical vs categorical features**

- **Visualisation using Python (Seaborn,Matplotlib) & Power BI.**
  Spatial KDE heatmap to explore hotspots.

---

2) Prediction approach: Supervised classification

- **Target variable: collision_severity**

- **Use historical patterns (spatial, temporal, road, environment) to predict accident severity**

---

3) Algorithms to be used

- **Random Forest – baseline, handles mixed features well**

- **Gradient Boosting – captures complex, non-linear risk patterns**

---

4) How results are used

- **Accident-prone zones:**
  Aggregate predicted severity probabilities spatially (lat-long / district level)

- **High-risk time windows:**
  Aggregate predictions by hour of day and day of week

---

5) Expected output

- **Map of accident hotspots**

- **Identification of high-risk time periods**

- **Decision support insights for road safety planning**