# CPE312 Capstone project

# Data Cleaning, Preprocessing, and Exploratory Data Analysis Checklist

## 1. การทำความสะอาดข้อมูลและการเตรียมข้อมูล (Data Cleaning and Preprocessing)

### 1.1 การประเมินคุณภาพข้อมูล (Data Quality Assessment)

|  |  |
| --- | --- |
| **Missing Values** ตรวจสอบขอบเขตของข้อมูลที่ขาดหายและวิธีการจัดการกับข้อมูลที่ขาดหาย | ✓ |
| **Outliers** ระบุค่าที่ผิดปกติและอธิบายวิธีการจัดการกับค่าผิดปกติเหล่านี้ (เช่น การลบ การปรับเปลี่ยน) | ✓ |
| **Duplicates** ตรวจสอบรายการข้อมูลที่ซ้ำกันและวิธีการจัดการกับข้อมูลซ้ำ | ✓ |
| **Data Types** ตรวจสอบความถูกต้องของประเภทข้อมูล (เช่น วันที่ ตัวเลข) | ✓ |
| อื่นๆ | □ |

### 1.2 Data Transformation

|  |  |
| --- | --- |
| **Normalization** ระบุว่ามีตัวแปรใดบ้างที่ถูกปรับสเกลหรือทำให้เป็นปกติและเหตุผลที่ทำ | ✓ |
| **การเข้ารหัส (Encoding)** ตรวจสอบตัวแปรเชิงหมวดหมู่ใดบ้างที่ถูกเข้ารหัสเพื่อใช้ในการวิเคราะห์ | ✓ |
| อื่นๆ | □ |

## 2. การวิเคราะห์ข้อมูลเบื้องต้น (EDA)

**2.1 สถิติเชิงพรรณนา**

|  |  |
| --- | --- |
| สรุปสถิติหลักสำหรับตัวแปรที่สำคัญ (ค่าเฉลี่ย มัธยฐาน โหมด ช่วง เป็นต้น) | ✓ |
| ระบุลักษณะหลักของข้อมูล (เช่น แนวโน้มกลาง การกระจายตัวของข้อมูล) | ✓ |
| อื่นๆ | □ |

**2.2 ข้อมูลเชิงลึกและรูปแบบเบื้องต้น (Initial Insights and Patterns)**

|  |  |
| --- | --- |
| ค้นหารูปแบบ แนวโน้ม หรือความสัมพันธ์ที่พบในข้อมูล | ✓ |
| ระบุข้อมูลที่สำคัญหรือข้อสังเกตที่น่าสนใจจากการวิเคราะห์เบื้องต้น | ✓ |
| อื่นๆ | □ |

**2.3 การแสดงข้อมูลด้วยกราฟ**

|  |  |
| --- | --- |
| จัดเตรียมการแสดงข้อมูลที่เกี่ยวข้อง (เช่น histograms, scatter plots, box plots) เพื่อสนับสนุนผลการค้นพบ | ✓ |
| กราฟแต่ละอันควรมีการระบุชื่อและอธิบายอย่างชัดเจนเพื่อความเข้าใจ | ✓ |
| อื่นๆ | □ |

## 3. สรุปผลจากการวิเคราะห์ข้อมูลเบื้องต้น (EDA)

|  |
| --- |
| สรุปข้อมูลเชิงลึกหลักที่ได้รับจากการวิเคราะห์เบื้องต้น  **แนวโน้มอุบัติเหตุตามช่วงเวลา**:   1. พบว่าอุบัติเหตุสูงสุดเกิดในช่วงเดือนเมษายน ซึ่งสอดคล้องกับเทศกาลสงกรานต์ที่มีการเดินทางจำนวนมาก 2. ช่วงปลายเดือนธันวาคมถึงต้นเดือนมกราคม ซึ่งตรงกับเทศกาลปีใหม่ ก็มีจำนวนอุบัติเหตุเพิ่มขึ้นเช่นกัน เนื่องจากการเดินทางและกิจกรรมเฉลิมฉลองในช่วงสิ้นปี   **5 อันดับสาเหตุการเกิดอุบัติเหตุเยอะสุดจากการ Visualization**  1.ขับรถเกินความเร็วกำหนด  2.คน/รถ/สัตว์ตัดหน้ากระชั้นชิด  3.หลับใน  4.เมาสุรา  5.อุปกรณ์ยานพาหนะบกพร่อง |