

```

from pyspark.sql.functions import col, when

df = df.withColumn(
    "weapon_category",
    when(col("weapon_description").contains("GUN"), "Firearm")
    .when(col("weapon_description").contains("RIFLE"), "Firearm")
    .when(col("weapon_description").contains("REVOLVER"), "Firearm")
    .when(col("weapon_description").contains("PISTOL"), "Firearm")
    .when(col("weapon_description").contains("FIREARM"), "Firearm")
    .when(col("weapon_description").contains("KNIFE"), "Knife/Sharp
Object")
    .when(col("weapon_description").contains("BLADE"), "Knife/Sharp
Object")
    .when(col("weapon_description").contains("RAZOR"), "Knife/Sharp
Object")
    .when(col("weapon_description").contains("SWORD"), "Knife/Sharp
Object")
    .when(col("weapon_description").contains("ICE PICK"), "Knife/Sharp
Object")
    .when(col("weapon_description").contains("CLEAVER"), "Knife/Sharp
Object")
    .when(col("weapon_description").contains("SCREWDRIVER"),
"Knife/Sharp Object")
    .when(col("weapon_description").contains("PIPE"), "Blunt Object")
    .when(col("weapon_description").contains("BAT"), "Blunt Object")
    .when(col("weapon_description").contains("HAMMER"), "Blunt
Object")
    .when(col("weapon_description").contains("BOARD"), "Blunt Object")
    .when(col("weapon_description").contains("IRON"), "Blunt Object")
    .when(col("weapon_description").contains("BRICK"), "Blunt Object")
    .when(col("weapon_description").contains("CHAIN"), "Blunt Object")
    .when(col("weapon_description").contains("BLUNT INSTRUMENT"),
"Blunt Object")
    .when(col("weapon_description").contains("MACE"),
"Chemical/Explosive")
    .when(col("weapon_description").contains("PEPPER SPRAY"),
"Chemical/Explosive")
    .when(col("weapon_description").contains("CHEMICAL"),
"Chemical/Explosive")
    .when(col("weapon_description").contains("EXPLOSIVE"),
"Chemical/Explosive")
    .when(col("weapon_description").contains("SCALDING LIQUID"),
"Chemical/Explosive")
    .when(col("weapon_description").contains("BOMB"),
"Chemical/Explosive")

```

```

        .when(col("weapon_description").contains("SIMULATED"), "Simulated
Weapon")
        .when(col("weapon_description").contains("TOY"), "Simulated
Weapon")
        .when(col("weapon_description").contains("STARTER PISTOL"),
"Simulated Weapon")
        .when(col("weapon_description").contains("DEMAND NOTE"),
"Verbal/Threat")
        .when(col("weapon_description").contains("VERBAL THREAT"),
"Verbal/Threat")
        .when(col("weapon_description").contains("DOG"),
"Biological/Animal/Body Force")
        .when(col("weapon_description").contains("PHYSICAL"),
"Biological/Animal/Body Force")
        .when(col("weapon_description").contains("STRONG-ARM"),
"Biological/Animal/Body Force")
        .when(col("weapon_description").contains("LIQUOR"),
"Biological/Animal/Body Force")
        .otherwise("Other/Unknown")
    )

```

+++++

```

from pyspark.sql.functions import col, when

```

```

df = df.withColumn(
    "crime_category",
    when(col("crime_description").contains("ASSAULT"), "Assault &
Battery")
    .when(col("crime_description").contains("BATTERY"), "Assault &
Battery")
    .when(col("crime_description").contains("HOMICIDE"), "Homicide &
Attempted Murder")
    .when(col("crime_description").contains("MURDER"), "Homicide &
Attempted Murder")
    .when(col("crime_description").contains("THEFT"), "Theft &
Burglary")
    .when(col("crime_description").contains("BURGLARY"), "Theft &
Burglary")
    .when(col("crime_description").contains("LARCENY"), "Theft &
Burglary")
    .when(col("crime_description").contains("ROBBERY"), "Robbery")
    .when(col("crime_description").contains("RAPE"), "Sex Crimes")

```

```

        .when(col("crime_description").contains("SEX"), "Sex Crimes")
        .when(col("crime_description").contains("DRUG"), "Drug-Related
Offenses")
        .when(col("crime_description").contains("WEAPON"), "Weapons
Offenses")
        .when(col("crime_description").contains("FIREARM"), "Weapons
Offenses")
        .when(col("crime_description").contains("CHILD"), "Child-Related
Crimes")
        .when(col("crime_description").contains("FRAUD"), "Fraud &
Forgery")
        .when(col("crime_description").contains("FORGERY"), "Fraud &
Forgery")
        .when(col("crime_description").contains("CREDIT"), "Fraud &
Forgery")
        .when(col("crime_description").contains("VEHICLE"), "Traffic &
Vehicle Offenses")
        .when(col("crime_description").contains("DRIVING"), "Traffic &
Vehicle Offenses")
        .when(col("crime_description").contains("LOITERING"), "Public
Disturbance & Nuisance")
        .when(col("crime_description").contains("DISORDERLY"), "Public
Disturbance & Nuisance")
        .when(col("crime_description").contains("DOMESTIC"), "Domestic &
Intimate Partner Violence")
        .when(col("crime_description").contains("PARTNER"), "Domestic &
Intimate Partner Violence")
        .when(col("crime_description").contains("PROSTITUTION"), "Sex Work
/ Human Trafficking")
        .when(col("crime_description").contains("PIMPING"), "Sex Work /
Human Trafficking")
        .otherwise("Other / Miscellaneous")
    )

```