Crime Data Analytics

Capstone Project Presentation by

Karan Jain- 210491

Objectives and Scope

- Clean and prepare district data (2001–2012).
- Apply exponential smoothing to forecast 2013–2020 crime figures.
- Validate forecasts against cleaned 2013–2014 data.
- Compute PCA, means & standard deviations, trend slopes.

Data Preparation

- Raw NCRB tables cleaned in Excel using Pivot Tables & Power Query.
- Exported for Python-based analysis.

STATE/UT	DISTRICT	YEAR	CrimeType	Count
A & N ISLANDS	ANDAMAN	2001	MURDER	13
A & N ISLANDS	ANDAMAN	2001	ATTEMPT TO MURDER	0
A & N ISLANDS	ANDAMAN	2001	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	0
A & N ISLANDS	ANDAMAN	2001	RAPE	3
A & N ISLANDS	ANDAMAN	2001	KIDNAPPING & ABDUCTION	2
A & N ISLANDS	ANDAMAN	2001	DACOITY	0
A & N ISLANDS	ANDAMAN	2001	ROBBERY	4
A & N ISLANDS	ANDAMAN	2001	BURGLARY	62
A & N ISLANDS	ANDAMAN	2001	THEFT	65
A & N ISLANDS	ANDAMAN	2001	RIOTS	13
A & N ISLANDS	ANDAMAN	2001	CRIMINAL BREACH OF TRUST	10
A & N ISLANDS	ANDAMAN	2001	CHEATING	8
A & N ISLANDS	ANDAMAN	2001	COUNTERFIETING	2
A & N ISLANDS	ANDAMAN	2001	ARSON	4
A & N ISLANDS	ANDAMAN	2001	HURT/GREVIOUS HURT	113
A & N ISLANDS	ANDAMAN	2001	DOWRY DEATHS	0
A & N ISLANDS	ANDAMAN	2001	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	18
A & N ISLANDS	ANDAMAN	2001	INSULT TO MODESTY OF WOMEN	1
A & N ISLANDS	ANDAMAN	2001	CRUELTY BY HUSBAND OR HIS RELATIVES	9
A & N ISLANDS	ANDAMAN	2001	IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES	0
A & N ISLANDS	ANDAMAN	2001	CAUSING DEATH BY NEGLIGENCE	0
A & N ISLANDS	ANDAMAN	2001	OTHER IPC CRIMES	310
A & N ISLANDS	ANDAMAN	2001	TOTAL IPC CRIMES	637
A & N ISLANDS	NICOBAR	2001	MURDER	0
A & N ISLANDS	NICOBAR	2001	ATTEMPT TO MURDER	0
A & N ISLANDS	NICOBAR	2001	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	0
A & N ISLANDS	NICOBAR	2001	RAPE	0
A & N ISLANDS	NICOBAR	2001	KIDNAPPING & ABDUCTION	0
A & N ISLANDS	NICOBAR	2001	DACOITY	0
A & N ISLANDS	NICOBAR	2001	ROBBERY	0
A & N ISLANDS	NICOBAR	2001	BURGLARY	2
A & N ISLANDS	NICOBAR	2001	THEFT	0
A & N ISLANDS	NICOBAR	2001	RIOTS	0
A & N ISLANDS	NICOBAR	2001	CRIMINAL BREACH OF TRUST	0
A & N ISLANDS	NICOBAR	2001	CHEATING	0
A & N ISLANDS	NICOBAR	2001	COUNTERFIETING	0
A & N ISLANDS	NICOBAR	2001	ARSON	0
A & N ISLANDS	NICOBAR	2001	HURT/GREVIOUS HURT	5
A & N ISLANDS	NICOBAR	2001	DOWRY DEATHS	0
A & N ISLANDS	NICOBAR	2001	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	1
A & N ISLANDS	NICOBAR	2001	INSULT TO MODESTY OF WOMEN	0
A & N ISLANDS	NICOBAR	2001	CRUELTY BY HUSBAND OR HIS RELATIVES	0
A & N ISLANDS	NICOBAR	2001	IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES	0

Methodology

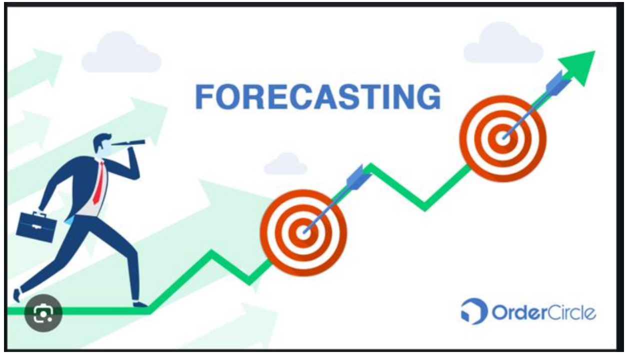
- Exponential Smoothing (Holt-Winters without seasonality):
- Smoothing parameters α (level) and β (trend) optimised via grid search.
- Implemented in Python using statsmodels library.

```python
import pandas as pd
from statsmodels.tsa.holtwinters import ExponentialSmoothing
import warnings

warnings.filterwarnings("ignore")

# ✅ Step 1: Load Excel File and NormalizedData sheet
file_path = "/kaggle/input/crime-data-district-wise/Crime_Data_District_Wise.xlsx"
sheet_name = "NormalizedData"

df = pd.read_excel(file_path, sheet_name=sheet_name)

# ✅ Step 2: Clean Data
df = df[['DISTRICT', 'CrimeType', 'YEAR', 'Count']]
df = df.dropna()
df = df[df['YEAR'] <= 2012]  # Use only historical data for forecasting

# ✅ Step 3: Forecast Setup
forecast_years = list(range(2013, 2021))  # Forecasting for 2013-2020
forecast_data = []

# ✅ Step 4: Forecasting loop per (District, CrimeType)
for (district, crime), group in df.groupby(['DISTRICT', 'CrimeType']):
    group = group.sort_values(by='YEAR')

    if group['Count'].sum() == 0 or len(group) < 3:
```

Validation (2013–2014)

- Mean Absolute Percentage Error (MAPE) ≈ 8.7%.
- Over 85% of districts achieved MAPE below 10%.
- Larger errors in districts with low or volatile crime counts.

Forecasts for 2015–2020



crime_forecast_2013_2020-2

DISTRICT	CrimeType	YEAR	Count
24 PARGANAS NORTH	ARSON	2013	32
24 PARGANAS NORTH	ARSON	2014	35
24 PARGANAS NORTH	ARSON	2015	38
24 PARGANAS NORTH	ARSON	2016	42
24 PARGANAS NORTH	ARSON	2017	45
24 PARGANAS NORTH	ARSON	2018	48
24 PARGANAS NORTH	ARSON	2019	51
24 PARGANAS NORTH	ARSON	2020	54
24 PARGANAS NORTH	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	2013	217
24 PARGANAS NORTH	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	2014	230
24 PARGANAS NORTH	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	2015	242
24 PARGANAS NORTH	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	2016	255
24 PARGANAS NORTH	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	2017	267
24 PARGANAS NORTH	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	2018	280
24 PARGANAS NORTH	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	2019	292
24 PARGANAS NORTH	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	2020	305
24 PARGANAS NORTH	ATTEMPT TO MURDER	2013	217
24 PARGANAS NORTH	ATTEMPT TO MURDER	2014	234
24 PARGANAS NORTH	ATTEMPT TO MURDER	2015	252
24 PARGANAS NORTH	ATTEMPT TO MURDER	2016	269
24 PARGANAS NORTH	ATTEMPT TO MURDER	2017	286
24 PARGANAS NORTH	ATTEMPT TO MURDER	2018	303
24 PARGANAS NORTH	ATTEMPT TO MURDER	2019	320
24 PARGANAS NORTH	ATTEMPT TO MURDER	2020	338
24 PARGANAS NORTH	BURGLARY	2013	14
24 PARGANAS NORTH	BURGLARY	2014	12

PCA Insights

- Reduced 20+ crime categories to 3 components (≈98% variance).
- PC1 (~86%): Overall crime volume.
- PC2 (~9%): Property vs. violent crime mix.
- PC3 (~3%): Niche crimes (arson, kidnapping).

pca_loadings

	PC1	PC2	PC3
ARSON_mean	0.18571516222233400	-0.18547638622055000	-0.410777305439517
ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY_mean	0.2313519784293790	-0.02544856915434770	-0.23074225546447100
ATTEMPT TO MURDER_mean	0.22570591166612700	-0.1684323570778130	0.08300679283606130
BURGLARY_mean	0.2257822934567350	0.3079784296044050	-0.0582745221925344
CAUSING DEATH BY NEGLIGENCE_mean	0.21370528071424300	-0.01894684827030150	-0.21622979337072800
CHEATING_mean	0.21627193240959000	0.26682348995553200	0.04796171386672040
COUNTERFIETING_mean	0.18581579922593100	0.3313506693015190	0.0960099724079664
CRIMINAL BREACH OF TRUST_mean	0.22074208508627000	0.17281635277429800	0.19581327713557600
CRUELTY BY HUSBAND OR HIS RELATIVES_mean	0.20320856800372400	-0.1913375868977720	-0.055433428632862100
CULPABLE HOMICIDE NOT AMOUNTING TO MURDER_mean	0.17865103863568900	-0.3126729814980510	0.31608836295399000
DACOITY_mean	0.1921862874497860	-0.1475212701381140	0.11364866188725900
DOWRY DEATHS_mean	0.20951940975117700	-0.27763478741897400	0.21452368050579700
HURT/GREVIOUS HURT_mean	0.2173790646328030	0.020343284363593100	-0.3402283055546490
IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES_mean	0.042705248465692400	-0.2536376614704450	0.3392612297279820
INSULT TO MODESTY OF WOMEN_mean	0.17092750347995000	0.11472312570742500	-0.239314667372666
KIDNAPPING & ABDUCTION_mean	0.21977424381388300	-0.05610171793873900	0.25715280160579800
MURDER_mean	0.24264456833791400	-0.1659347725321430	0.09529681408275390
OTHER IPC CRIMES_mean	0.23065202553847400	-0.0031362076265708100	-0.11610298564531900
RAPE_mean	0.23248263394977200	-0.12889957896451300	-0.024341078548382400
RIOTS_mean	0.18188591581749900	-0.26366662342371200	-0.1971359370889010
ROBBERY_mean	0.20757697042254100	0.2602581484420660	0.18269122823679700
THEFT_mean	0.21352877499081800	0.3538630746950490	0.2387196560947800
TOTAL IPC CRIMES_mean	0.2565909742112370	0.09804418552188720	-0.05063845547833780

Trend Slope

- Calculated per district for 2001–2012 using least-squares regression.
- Represents average annual increase or decrease in crime incidents.
- Helps identify rising, stable, or improving districts.

crime_trends

DISTRICT	Trend_Slope
THRISSUR RURAL	2969.7
CHENNAISUBURBAN	2606.5000000000000
SILIGURI_PC	2152.000000000000
TOTAL	2132.301555012380
OUTER	1534.1428571428600
CHENGAI	1476.7
TRIVANDRUM	1322.0000000000000
KOLKATA	1286.5054945054900
ERNAKULAM COMMR.	1220.1223776223800
ERNAKULAM RURAL	1191.1608391608400
SOUTH-EAST	1183.785714285710
MURSHIDABAD	1065.3252747252700
24 PARGANAS SOUTH	989.5659340659340
SURAT COMMR.	988.8516483516480
KOLLAM COMMR.	958.7000000000000
CYBERABAD	892.286713286713
MALAPPURAM	870.123076923077
NORTH-EAST	860.430303030303
ERNAKULAM	830.0000000000000
JAIPUR WEST	810.6
MUMBAI	810.5
NADIA	808.3472527472530
MUMBAI COMMR.	785.0333333333340
24 PARGANAS NORTH	772.9120879120880
CHENNAI	710.2175824175820
PATNA	707.9340659340660

District Wise Statistics

DISTRICT	ARSON_mean	ARSON_std	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY_mean	ASSAULT ON WOM
24 PARGANAS NORTH	13.0	13.638181696985900	171.23076923076900	
24 PARGANAS SOUTH	11.538461538461500	6.552548926631330	126.92307692307700	
ADILABAD	33.0	6.883648406522180	159.0	
AGAR	12.0	0.0	48.5	
AGRA	9.714285714285710	4.427685129319340	128.42857142857100	
AHMEDABAD COMMR.	7.923076923076920	6.70151150191182	76.92307692307690	
AHMEDABAD RURAL	14.5	17.403580533459600	22.642857142857100	
AHMEDNAGAR	43.07142857142860	7.518840803805270	127.85714285714300	
AHWA-DANG	1.3846153846153800	1.192927878405450	2.769230769230770	
AIZAWL	10.5	6.83599072781475	32.642857142857100	
AJMER	15.928571428571400	8.042811273852770	97.85714285714290	
AKOLA	34.714285714285700	6.329765719671940	109.14285714285700	
ALAPUZHA	21.0	7.942098153317110	194.57142857142900	
ALIGARH	30.5	21.306373188826400	114.35714285714300	
ALIRAJPUR	11.0	1.6329931618554500	23.571428571428600	
ALLAHABAD	13.928571428571400	11.67814715094480	84.92857142857140	
ALMORA	2.2857142857142900	2.0164161440372500	4.857142857142860	
ALWAR	27.285714285714300	10.64409241301030	144.5	
AMBALA	7.636363636363640	2.3354968324845700	21.363636363636400	
AMBALA RURAL	3.5	0.7071067811865480	22.0	
AMBALA URBAN	5.5	0.7071067811865480	35.5	
AMBEDKAR NAGAR	4.0	4.69041575982343	14.642857142857100	
AMETHI	7.0	9.899494936611670	80.5	
AMRAVATI COMMR.	13.142857142857100	4.521013209399020	51.142857142857100	
AMRAVATI RURAL	60.92857142857140	16.438530187497700	164.07142857142900	
AMRELI	9.142857142857140	5.332875438219090	22.071428571428600	
AMRITSAR	3.3	2.5407785333546000	12.9	

INTENT TO OUTRAGE HER MODESTY_std	ATTEMPT TO MURDER_mean	ATTEMPT TO MURDER_std	BURGLARY_mean	BURGLARY_std	CAUSING DEAT
63.479857495841200	118.61538461538500	94.15903077731350	30.615384615384600	20.365569234774900	
50.928809035197300	15.76923076923080	33.374525829724900	53.07692307692310	75.65542670386300	
38.99112325014110	72.85714285714290	16.209480335509000	220.35714285714300	33.48388163983280	
10.606601717798200	24.0	9.899494936611670	43.5	14.849242404917500	
65.46721350382990	151.07142857142900	29.155984438386400	266.5	91.14548809458420	
35.46468087750950	65.07692307692310	24.841033051725600	774.3076923076920	97.81648175314890	
6.957089829885690	20.714285714285700	14.274391116900300	104.92857142857100	21.225761871673000	
62.76801873425810	68.57142857142860	22.527638702147800	454.2142857142860	110.75973898587800	
2.27866357593825	1.0769230769230800	0.8623164985025760	4.615384615384620	2.754948926884780	
12.707400284008600	10.571428571428600	5.214097542893990	197.28571428571400	58.03371509658920	
36.6728932442083	53.642857142857100	13.669770933071200	291.92857142857100	97.91869805390300	
44.02421910871180	48.07142857142860	12.356650565052500	266.14285714285700	47.571577071345300	
50.28949161637020	28.857142857142900	7.862835100914670	196.92857142857100	31.059407804831300	
93.62495323433740	199.14285714285700	39.99587890858670	178.0	72.97523289023900	
8.202786982831770	37.0	10.98483803552270	36.714285714285700	10.656989658033300	
81.19157239868820	121.07142857142900	40.53766394110790	257.0	67.30641757034650	
1.4064216928154900	6.571428571428570	4.1826433620983600	16.571428571428600	7.408044335661130	
55.844220140287800	80.28571428571430	22.331173696904500	222.5	46.28797817011370	
5.536654716933810	19.90909090909090	8.166450213581730	269.0	61.72195719515060	
14.142135623731000	10.0	7.0710678118654800	105.0	18.384776310850200	
17.67766952966370	15.5	0.7071067811865480	378.0	22.627416997969500	
6.934940994226740	32.42857142857140	10.248559266207500	39.5	20.19043947401460	
3.5355339059327400	55.5	7.7781745930520200	56.0	7.0710678118654800	
28.530300068138500	29.714285714285700	9.611119933370220	232.92857142857100	35.60783035257250	
89.72561960324830	38.142857142857100	12.859340471526400	268.0	58.301603091826900	

Key Findings

- Simple exponential smoothing delivers reliable forecasts.
- Forecasts enable proactive resource allocation.
- PCA and trend slopes offer strategic insights.

References

- 1. NCRB. Crime in India Reports (2001–2012).
- 2. Jolliffe, I. T. (2002). Principal Component Analysis.
- 3. Hotelling, H. (1933). Analysis of a Complex of Statistical Variables.
- 4. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting: Methods and Applications.

Thank You

Special thanks to my supervisor Prof. Deepu Philip for their guidance and support throughout this project.