📦 Importing necessary libraries

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

🎨 Set seaborn style for clean visuals

In [5]:
```
sns.set(style="whitegrid")
```

📂 Step 1: Load the cleaned dataset

In [8]:
```
df = pd.read_csv("tn_crime_data_2014_2021_final_cleaned.csv")
```

In [10]:
```
df
```

Out[10]:

	District	2014	2015	2016	2017	2018	2019	2020	2021
0	0	3319	74049	151417	228785	306153	383521	460889	538257
1	Ariyalur	3432	2098	763	571	1906	3240	4575	5909
2	Chengalpattu	6770	2732	1306	5344	9382	13420	17458	21496
3	Chennai	33049	19065	5080	8904	22889	36873	50858	64842
4	Coimbatore	1802	734	3269	5805	8340	10876	13411	15947
5	Coimbatore City	1959	2545	3130	3716	4301	4887	5472	6058
6	Cuddalore	8790	3785	1220	6225	11230	16235	21240	26245
7	CyberCell	2	2	3	4	4	4	5	6
8	Dharmapuri	741	732	2204	3676	5149	6622	8094	9566
9	Dindigul	5891	1894	2104	6102	10099	14096	18094	22092
10	Erode	6790	6584	6378	6172	5966	5760	5554	5348
11	Kallakurichi	5770	3187	604	1979	4562	7145	9728	12311
12	Kanchipuram	22047	20090	18134	16178	14221	12264	10308	8352
13	Kanyakumari	6270	5989	5707	5426	5144	4863	4581	4300
14	Karur	9185	6760	4336	1912	513	2938	5362	7786
15	Krishnagiri	7152	7427	7703	7978	8254	8529	8805	9080
16	Madurai	14211	15472	16734	17996	19257	20518	21780	23042
17	Madurai City	837	2149	3461	4773	6085	7397	8709	10021
18	Nagapattinam	16890	11815	6740	1665	3410	8485	13560	18635
19	Namakkal	2617	1202	214	1630	3045	4460	5876	7292
20	Nilgiris	2597	2804	3010	3216	3423	3630	3836	4042
21	Other Units	0	0	0	0	0	0	0	0

	District	2014	2015	2016	2017	2018	2019	2020	2021
22	Perambalur	3328	2053	778	497	1772	3047	4322	5597
23	Pudukottai	11060	7867	4674	1481	1712	4905	8098	11291
24	Railway Chennai	11115	9634	8152	6670	5189	3708	2226	744
25	Railway Trichy	1626	1408	1190	972	754	536	318	100
26	Ramnathapuram	563	1115	1667	2219	2771	3323	3875	4427
27	Ranipet	7808	5897	3986	2075	164	1747	3658	5569
28	Salem	4544	4956	5369	5781	6194	6606	7019	7431
29	Salem City	4802	4894	4986	5078	5170	5262	5354	5446
30	Sivagangai	1165	2034	2903	3772	4641	5510	6379	7248
31	Tenkasi	10254	7946	5638	3330	1022	1286	3594	5902
32	Thanjavur	28118	19440	10762	2084	6594	15272	23950	32628
33	Theni	2809	182	3174	6166	9157	12148	15140	18132
34	Thirunelveli	17605	16027	14449	12871	11293	9715	8137	6559
35	Thirunelveli City	1592	1887	2181	2476	2770	3065	3359	3654
36	Thiruvallur	20389	20301	20213	20125	20037	19949	19861	19773
37	Thiruvannamalai	8003	12230	16458	20685	24913	29140	33368	37595
38	Thiruvarur	3335	3400	3464	3528	3593	3658	3722	3786
39	Thoothukudi	3201	4029	4856	5684	6511	7339	8166	8994
40	Tiruppattur	7935	6262	4589	2916	1243	430	2103	3776
41	Tiruppur	11074	6805	2537	1732	6000	10269	14537	18806
42	Tiruppur City	12709	9581	6453	3325	197	2931	6059	9187
43	Trichy	4163	1255	1652	4560	7467	10375	13282	16190

	District	2014	2015	2016	2017	2018	2019	2020	2021
44	Trichy City	3432	3540	3649	3758	3866	3974	4083	4192
45	Vellore	27643	24758	21873	18988	16103	13218	10333	7448
46	Villupuram	12491	11434	10377	9320	8263	7206	6149	5092
47	Virudhunagar	7505	7336	7167	6998	6829	6660	6491	6322

🧹 Step 2: Prepare the data for analysis

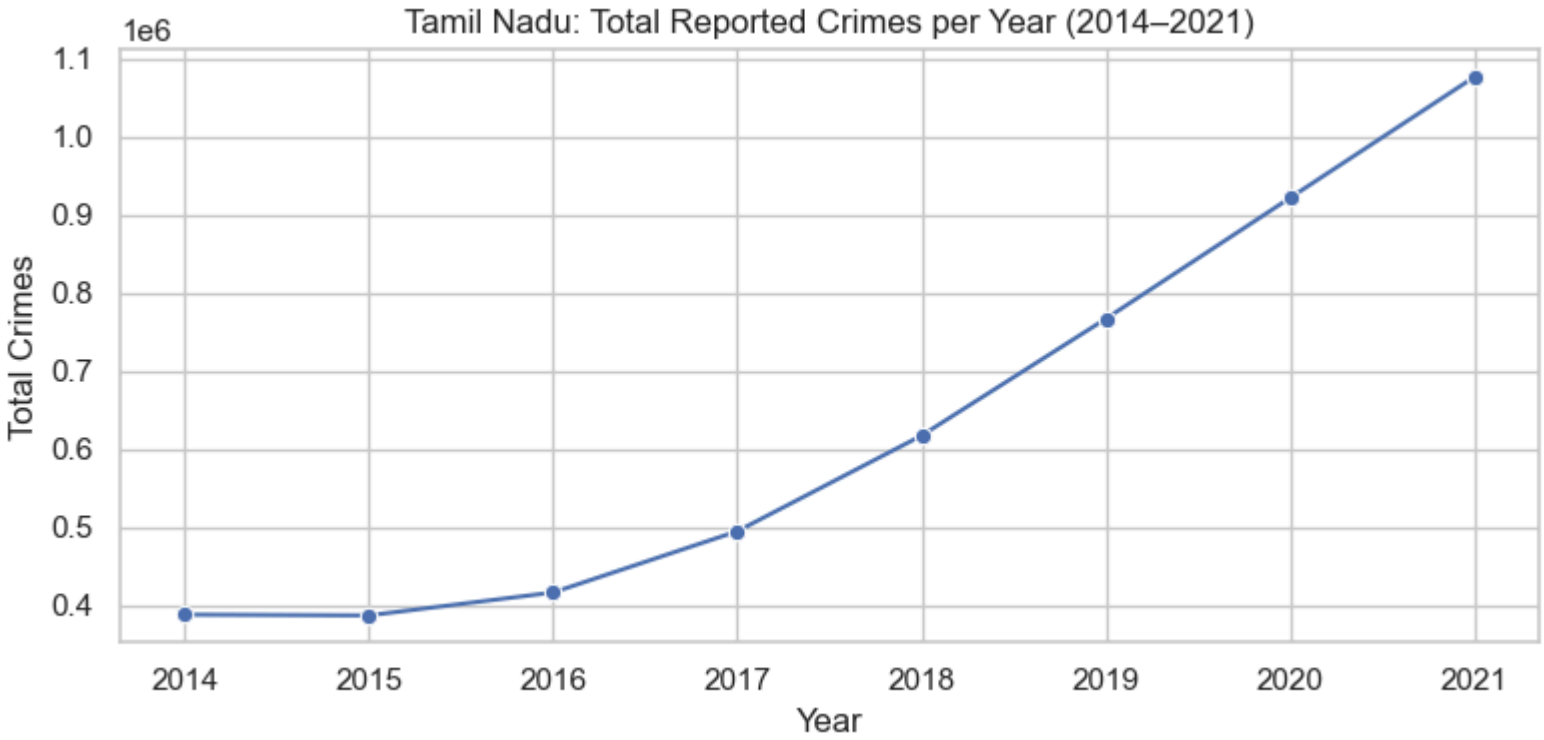
We'll reshape the data into long format for year-wise analysis

In [13]:
```
year_cols = [str(year) for year in range(2014, 2022)]
df_long = df.melt(id_vars="District", value_vars=year_cols,
                  var_name="Year", value_name="Crime_Count")
df_long["Year"] = df_long["Year"].astype(int)
```

📈 Trend Analysis: Total Crimes per Year (Statewide)

In [20]:
```
plt.figure(figsize=(8, 4))
sns.lineplot(
    data=df_long,
    x="Year",
    y="Crime_Count",
    estimator="sum",
```

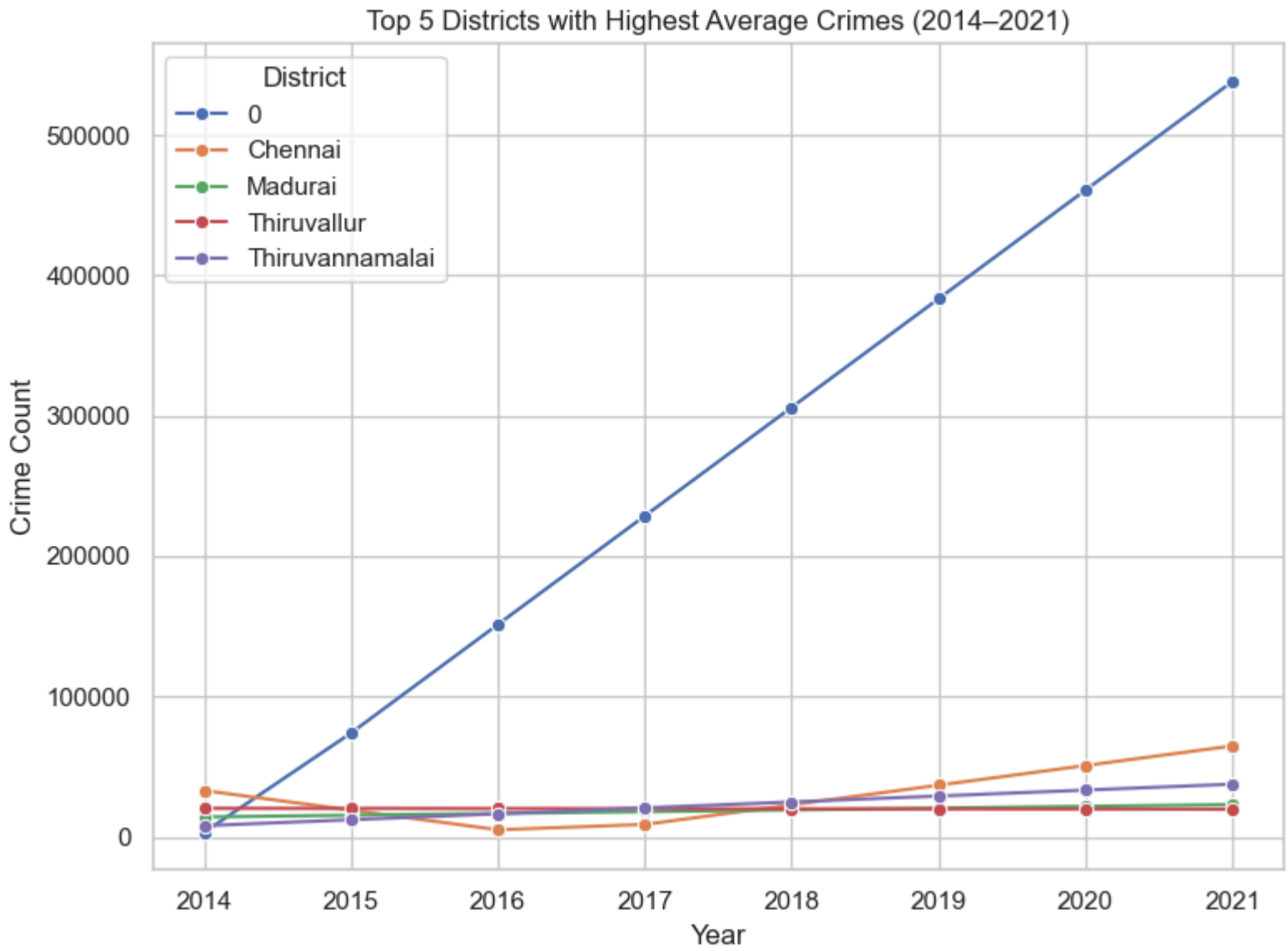
```
    errorbar=None,
    marker="o"
)
plt.title("Tamil Nadu: Total Reported Crimes per Year (2014–2021)")  # 👉 No emoji
plt.xlabel("Year")
plt.ylabel("Total Crimes")
plt.grid(True)
plt.tight_layout()
plt.savefig("trend_total_crimes.png")
plt.show()
```



🏙️ Identify Top 5 High-Crime Districts (on avg)

Let's find districts with highest average crime over 8 years

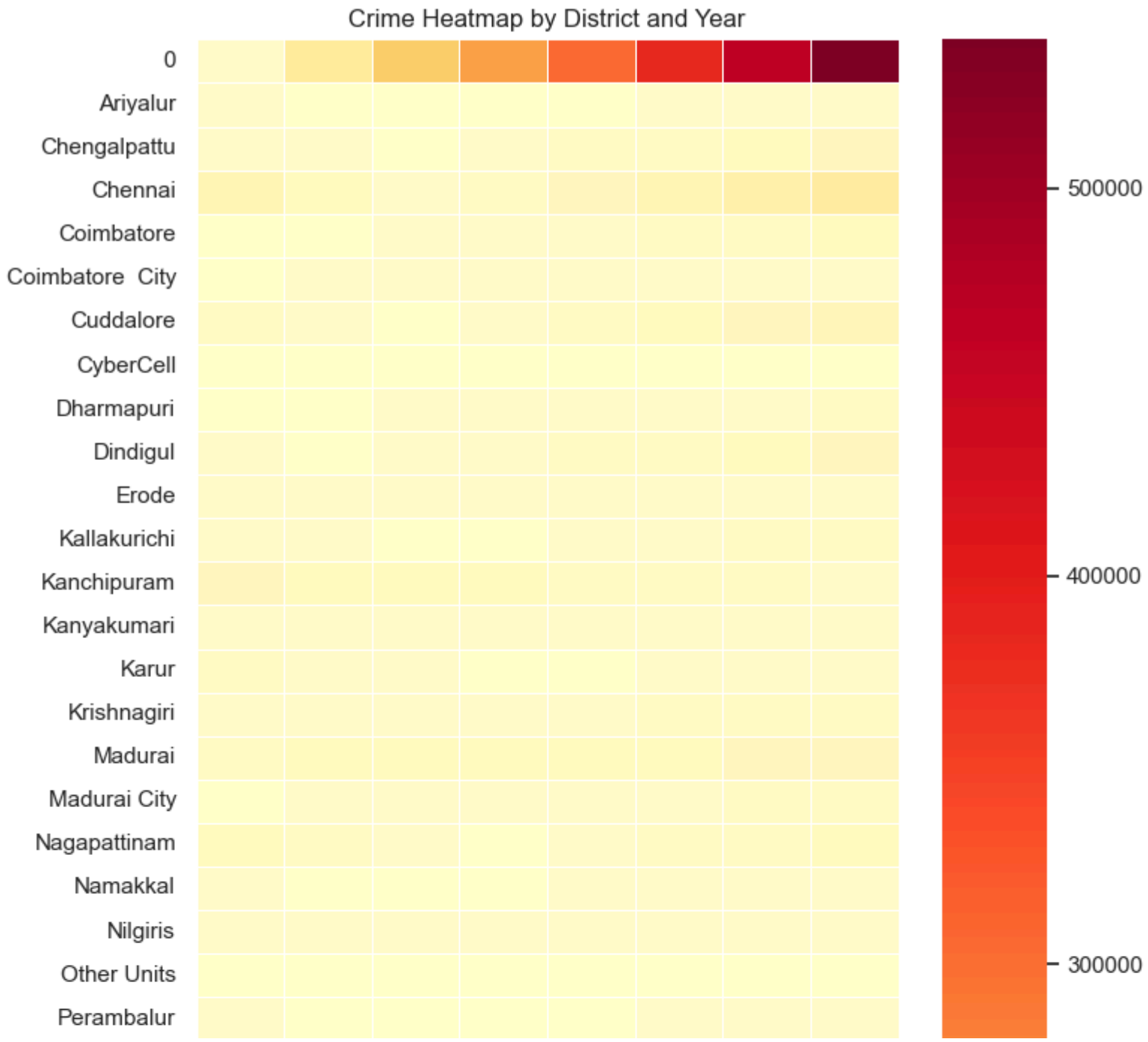
In [30]:
```
avg_crimes = df_long.groupby("District")["Crime_Count"].mean().sort_values(ascending=False)
top_districts = avg_crimes.head(5).index.tolist()


df_top = df_long[df_long["District"].isin(top_districts)]


plt.figure(figsize=(8, 6))
sns.lineplot(data=df_top, x="Year", y="Crime_Count", hue="District", marker="o")
plt.title("Top 5 Districts with Highest Average Crimes (2014-2021)")
plt.xlabel("Year")
plt.ylabel("Crime Count")
plt.legend(title="District")
plt.grid(True)
plt.tight_layout()
plt.savefig("top5_districts_trend.png")
plt.show()
```

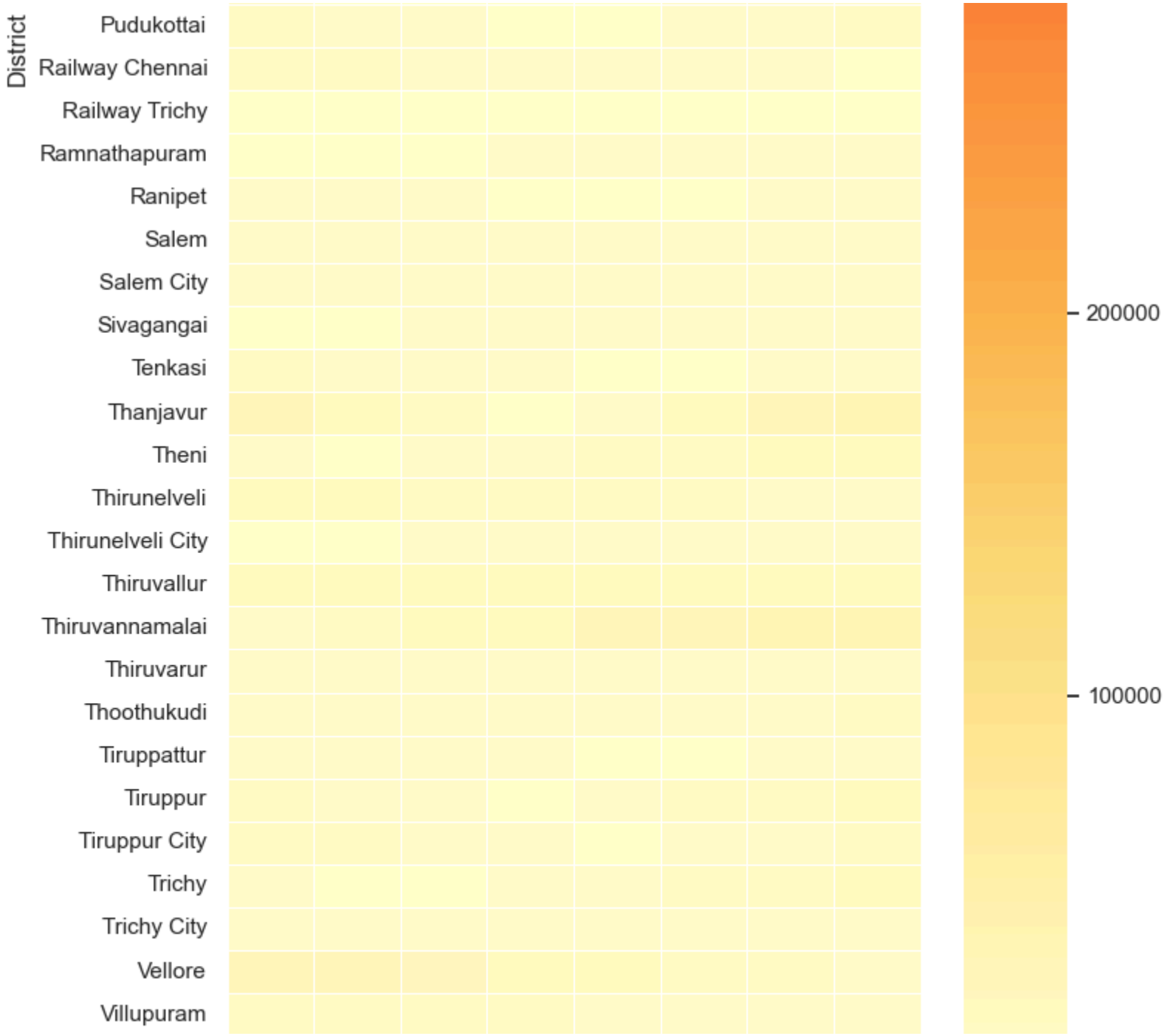

Top 5 Districts with Highest Average Crimes (2014–2021)
District
0
Chennai
Madurai
Thiruvallur
Thiruvannamalai
Crime Count
500000
400000
300000
200000
100000
0
2014
2015
2016
2017
2018
2019
2020
2021
Year

🔥 Crime Intensity Heatmap by District and Year

A heatmap gives a clear picture of crime distribution over time

In [34]:
```
crime_matrix = df.set_index("District")[year_cols]

plt.figure(figsize=(8, 15))
sns.heatmap(crime_matrix, cmap="YlOrRd", linewidths=0.5)
plt.title("Crime Heatmap by District and Year")
plt.xlabel("Year")
plt.ylabel("District")
plt.tight_layout()
plt.savefig("crime_heatmap_years.png")
plt.show()
```


Crime Heatmap by District and Year
0
Ariyalur
Chengalpattu
Chennai
Coimbatore
Coimbatore City
Cuddalore
CyberCell
Dharmapuri
Dindigul
Erode
Kallakurichi
Kanchipuram
Kanyakumari
Karur
Krishnagiri
Madurai
Madurai City
Nagapattinam
Namakkal
Nilgiris
Other Units
Perambalur
500000
400000
300000


District
Pudukottai
Railway Chennai
Railway Trichy
Ramnathapuram
Ranipet
Salem
Salem City
Sivagangai
Tenkasi
Thanjavur
Theni
Thirunelveli
Thirunelveli City
Thiruvallur
Thiruvannamalai
Thiruvarur
Thoothukudi
Tiruppattur
Tiruppur
Tiruppur City
Trichy
Trichy City
Vellore
Villupuram
200000
100000

📊 Correlation: Which Districts Follow the State Trend?

We'll compute correlation between each district's trend and total TN trend

In [ ]:
```
df["Total_Crime"] = df[year_cols].sum(axis=1)
```

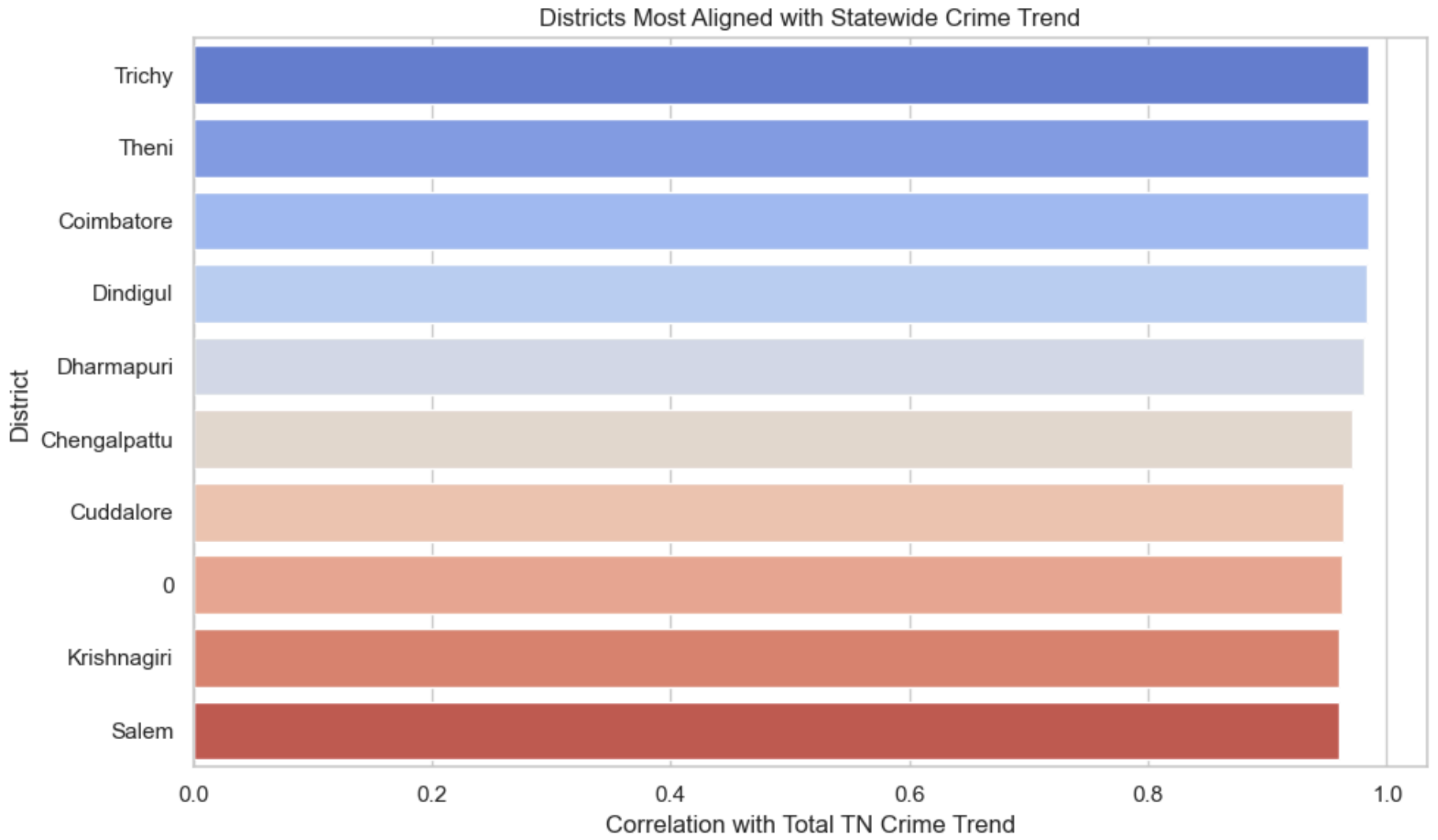
Transpose to make year-wise correlation possible

In [40]:
```
transposed = df.set_index("District")[year_cols].T

total_crime_trend = transposed.sum(axis=1)

valid_cols = transposed.loc[:, transposed.std() != 0].dropna(axis=1, how='any')

correlation_scores = valid_cols.corrwith(total_crime_trend).sort_values(ascending=False)


print(correlation_scores.head(10))
```

```
District
Trichy          0.984670
Theni           0.984088
Coimbatore      0.984069
Dindigul        0.983267
Dharmapuri      0.980338
Chengalpattu    0.971157
Cuddalore       0.963680
0               0.962545
Krishnagiri     0.960388
Salem           0.960386
dtype: float64
```

Top 10 most trend-aligned districts

In [45]:
```
top_corr = correlation_scores.head(10).reset_index()
top_corr.columns = ['District', 'Correlation']

plt.figure(figsize=(10, 6))
sns.barplot(
    data=top_corr,
    x="Correlation",
    y="District",
    hue="District",
    palette="coolwarm",
    dodge=False,
    legend=False
)
plt.title("Districts Most Aligned with Statewide Crime Trend")
plt.xlabel("Correlation with Total TN Crime Trend")
plt.ylabel("District")
plt.tight_layout()
plt.savefig("correlated_districts.png")
plt.show()
```


Districts Most Aligned with Statewide Crime Trend
Trichy
Theni
Coimbatore
Dindigul
Dharmapuri
Chengalpattu
Cuddalore
0
Krishnagiri
Salem
District
0.0
0.2
0.4
0.6
0.8
1.0
Correlation with Total TN Crime Trend

In [ ]: