Crimes em Portugal: Análise de Dados para uma Empresa de Segurança

Índice:

* Introdução:
  + Dados
  + Lista de Perguntas
* Análise de Dados:
  + Processamento de Dados
  + Análise Exploratória de Dados
  + Criação de Visualizações
* Resultados
* Conclusões

Introdução:

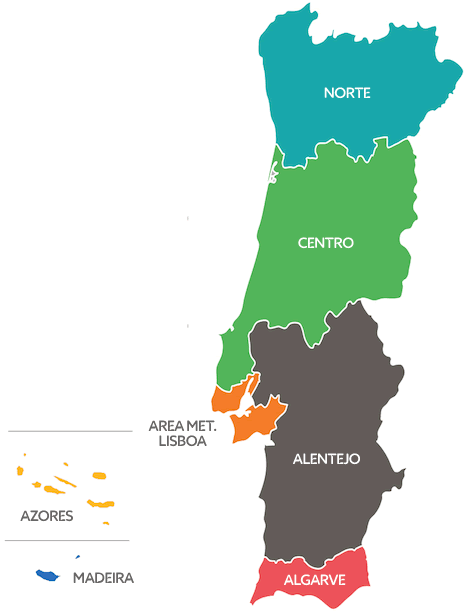
 Num mundo cada vez mais interligado, a segurança tornou-se uma prioridade máxima tanto para indivíduos quanto para empresas. Para uma empresa na indústria de alarmes de segurança, obter conhecimento sobre o panorama criminal local é crucial. Este projeto é impulsionado pela necessidade de uma empresa de alarmes de segurança compreender a situação criminal em Portugal, à medida que estabelece a sua presença neste país mediterrâneo. A visão da empresa abrange três cidades principais - Lisboa, Porto e Algarve - onde pretendem abrir escritórios. No entanto, questionam se esses são os locais ideais para abrir novos escritórios?

Figura 1 – Regiões de Portugal

**Dados**

Para atingir este objetivo, utilizei um conjunto de dados do Kaggle com o registo de crimes em Portugal, por diferentes localizações, de 2011 a 2019.

O conjunto de dados, proveniente do Kaggle, oferece uma visão abrangente das estatísticas criminais em Portugal, com 166 colunas e 312 linhas, proporcionando uma riqueza de pontos de dados que pintam um quadro detalhado dos crimes em diferentes regiões (figure 1). As colunas incluem Território, Região, Concelho, Município e as colunas subsequentes representam o número de crimes cometidos por tipo e ano (figure 2). As primeiras quatro colunas são dados do tipo texto e as restantes são numéricas.

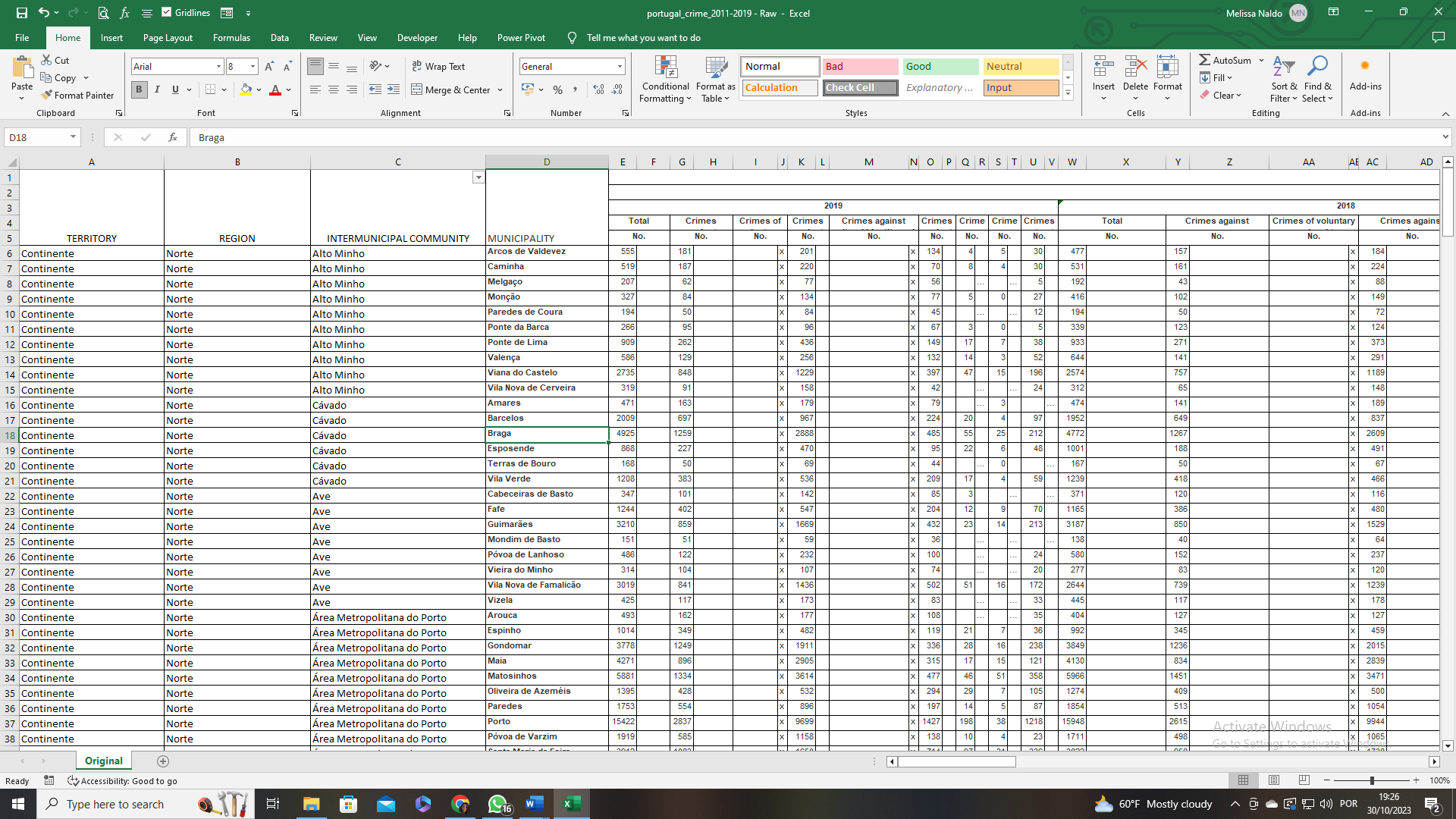


Figura 2 – Dados originais.

O conjunto de dados pode ser acessado através do seguinte link: https://www.kaggle.com/datasets/sameerkulkarni91/portugal-crime-statistics-20112019?select=portugal\_consolidated\_2011-2019.xlsx

**Lista de Perguntas:**

Esta análise é guiada por um conjunto de perguntas pertinentes que visam descobrir padrões subjacentes nos dados criminais e informar as decisões estratégicas da empresa:

1. Quantos crimes foram registados em Portugal de 2011 a 2019?
2. Qual é a extensão dos crimes contra o património dentro do mesmo período?
3. Como varia a criminalidade entre as regiões e qual é a média de criminalidade em cada região?
4. Como evoluiu a criminalidade ao longo dos anos?
5. Quais são os 10 municípios com os maiores valores de criminalidade?
6. Quais são os 3 principais locais para abrir os escritórios que seriam mais benéficos para a empresa?

**Análise de Dados:**

**Pré-processamento de Dados:**

Para uma análise inicial do conjunto de dados, várias etapas de processamento foram realizadas:

***Remoção de colunas e linhas:***

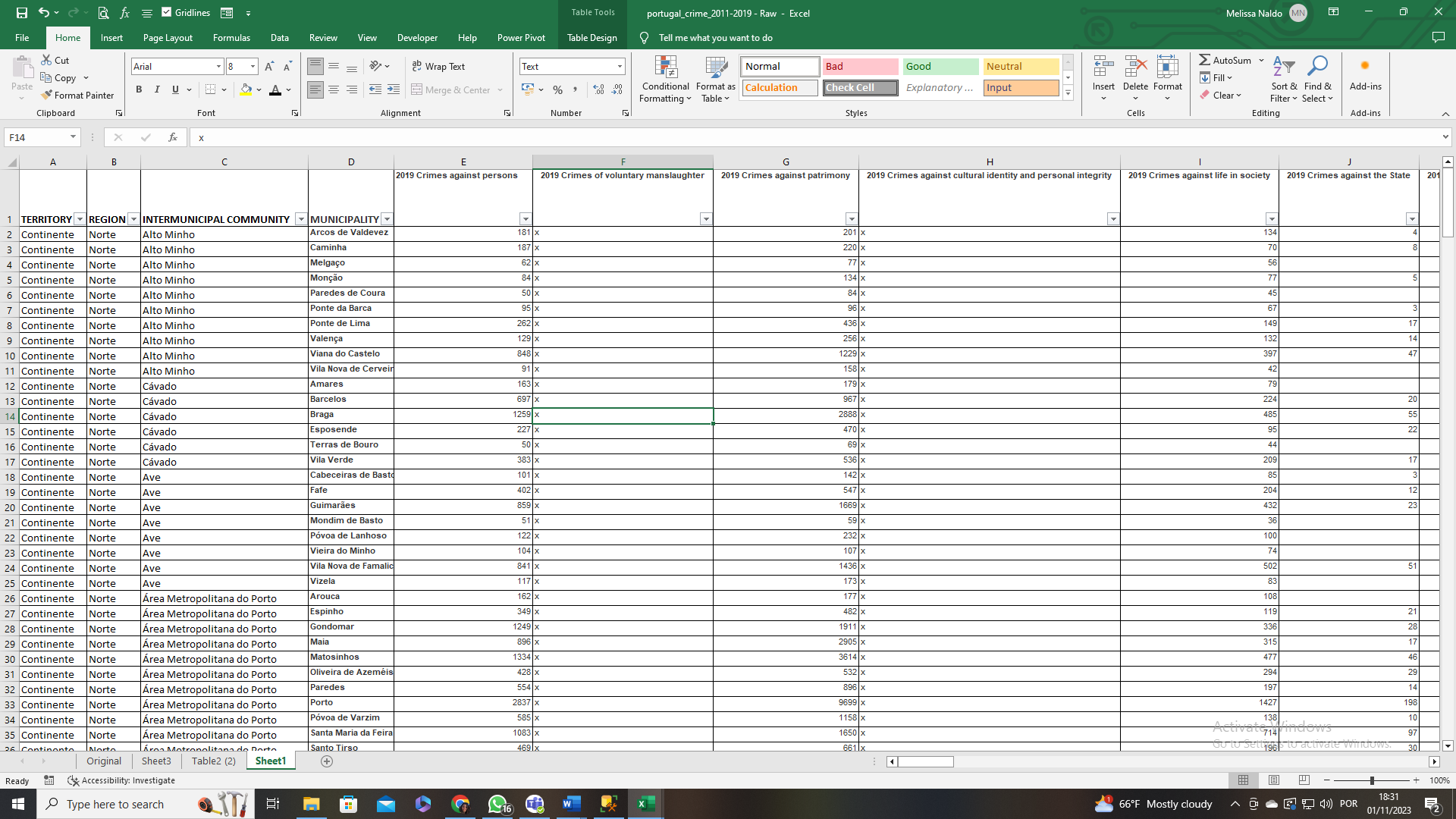
Nenhuma linha duplicada foi identificada no conjunto de dados. Colunas sem informações significativas, marcadas com “x” ou “-”, foram removidas. A coluna “Total” também foi eliminada dado não contribuir para a análise. Posteriormente, os dados foram transformados combinando a coluna dos anos com o tipo de crime, criando um único cabeçalho para estas colunas (figura 3).

Figura 3 – Dados com apenas um cabeçálho.

***Desnormalização de Colunas:***

O Editor do Power Query foi utilizado para ajudar a descentralizar as colunas, resultando num conjunto de dados mais estruturado. A seguir, usei as funções Left(), Len() e Right() para separar os anos dos tipos de crimes. Com isso, criei colunas distintas para tipos e anos de crimes, levando à seguinte imagem (figura 4):

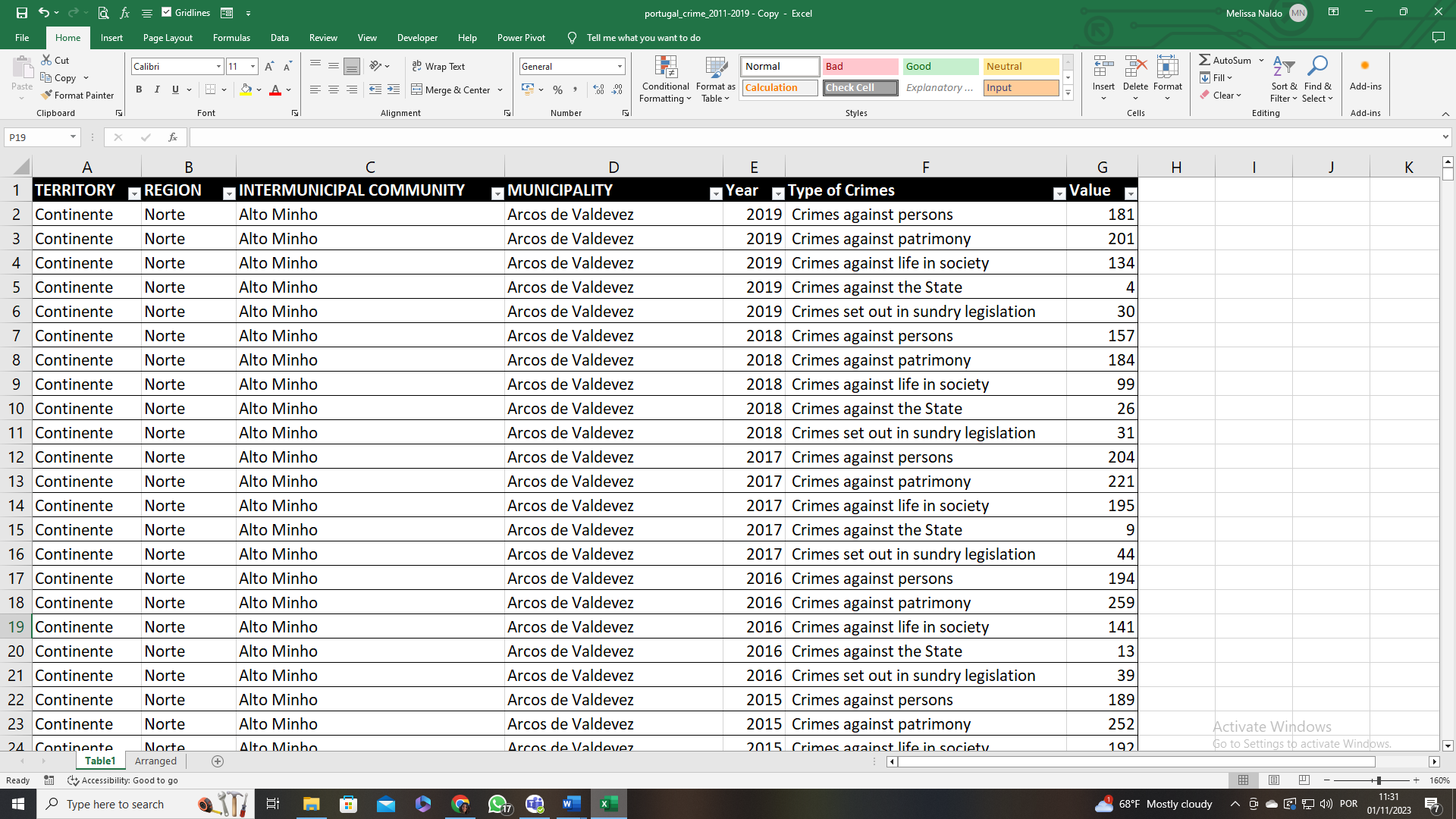


Figura 4 – Dados depois de criar colunas diferentes.

***Padronização de Dados:***

Para garantir a uniformidade dos dados, os valores do texto foram padronizados. Além disso, erros ortográficos foram corrigidos e alguns valores do texto foram alterados para aumentar a precisão (figura 5):

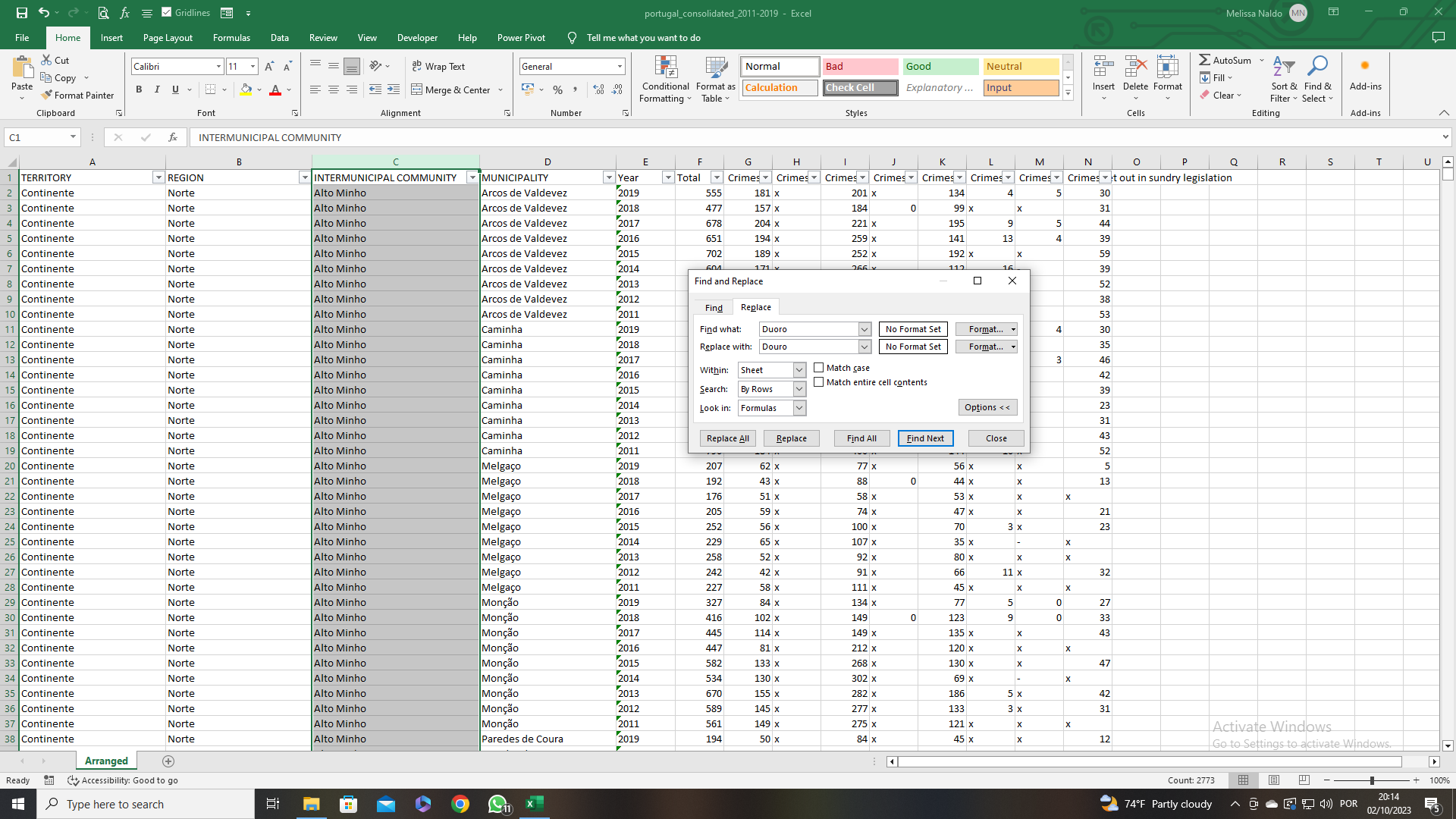
* "Região Autónoma da Madeira" para "RA Madeira";
* "Região Autónoma dos Açores" para "RA Açores";
* "Área Metropolitana de Lisboa" para "AM Lisboa";
* "Área Metropolitana do Porto" para "AM Porto";
* "Duoro" para "Douro".

Figura 5 - Recurso Localizar e Substituir.

***Tratamento de Valores Ausentes:***

Vários “Tipos de Crime” tinham valores ausentes. Essas categorias eram “voluntary manslaughter”, “cultural identity & personal integrity”, “pet animals”, “life in society”, “the State” e “set out in sundry legislation”. As três primeiras colunas tinham muitos dados em falta e foram removidas porque não contribuíam significativamente para os objetivos do projeto. Para as demais colunas, os valores ausentes foram substituídos pelos valores médios, que eram semelhantes à moda e à mediana, e esses valores foram: 148 para “crime against life in society”, 26 para “crimes against the State” e 96 para " crimes set out in sundry legislation ".

***Transformação dos nomes das colunas:***

Com isto, finalizei a parte da análise, correspondendo à da limpeza dos dados, ao alterar os nomes das minhas colunas (figura 6).

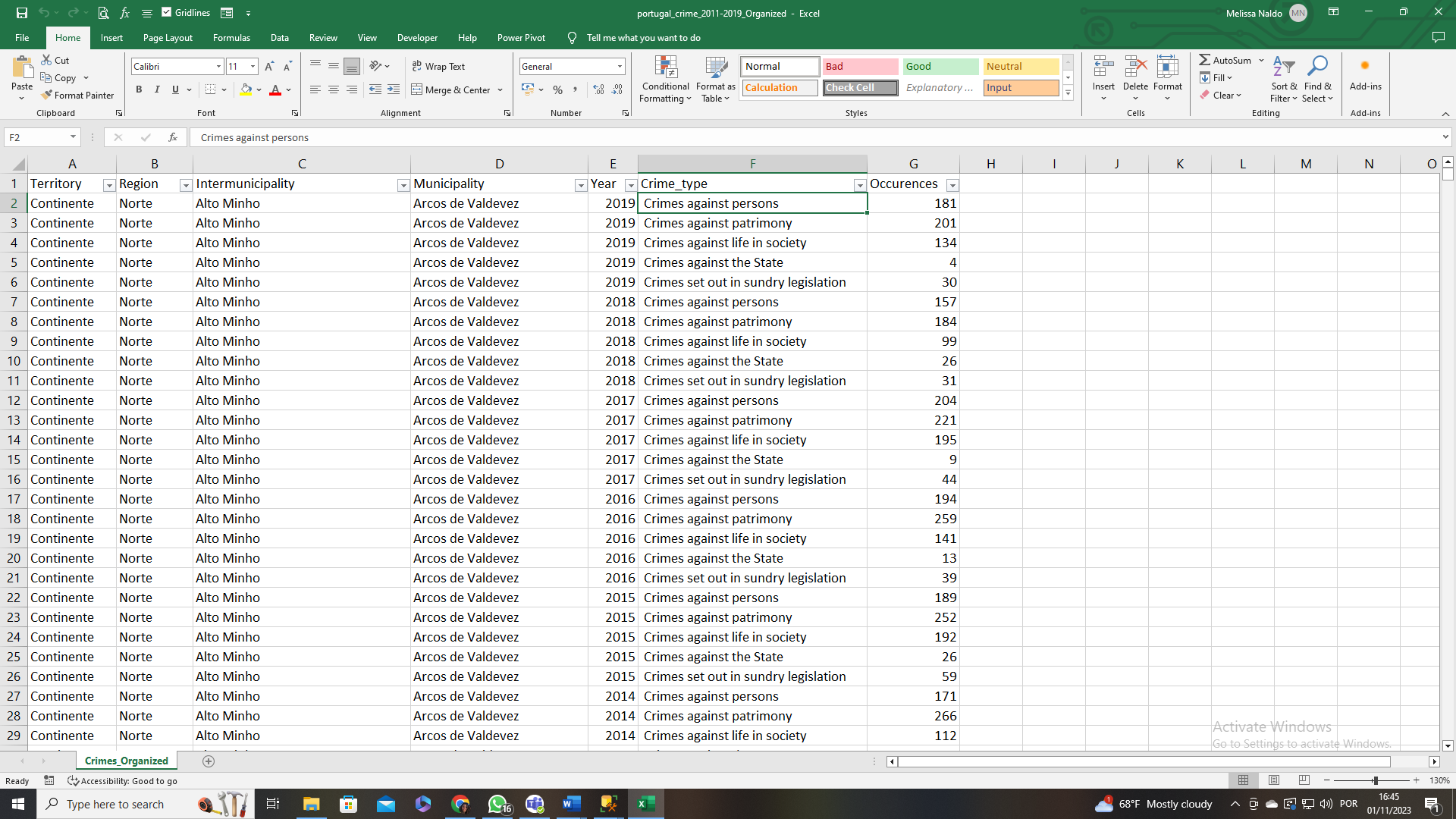


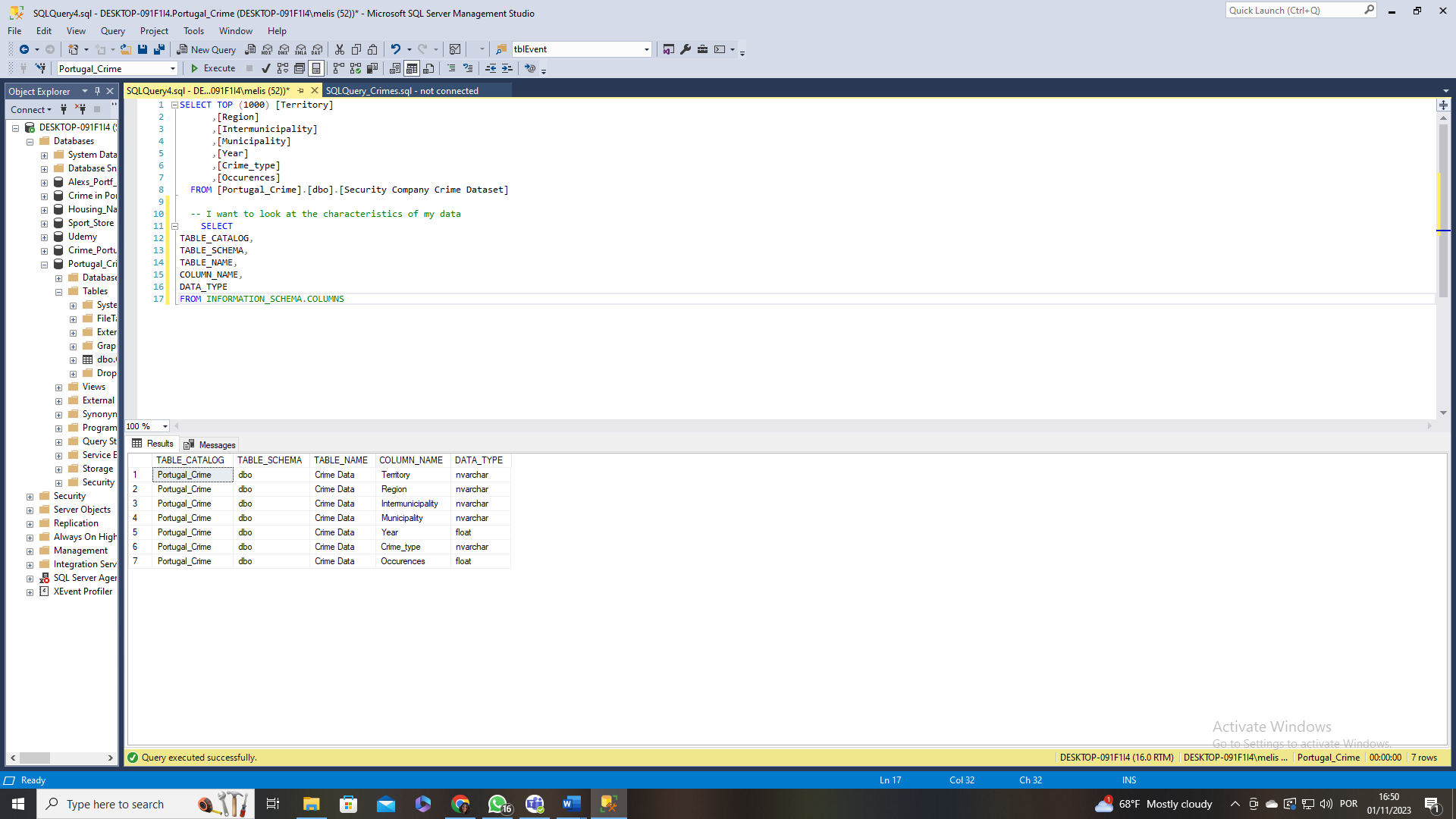
Figura 6- Dados limpos.

Pode aceder ao conjunto de dados iniciais através do link (https://github.com/Melissa-Naldo/Crime\_In\_Portugal/blob/main/portugal\_crime\_2011-2019%20-%20Raw.xlsx) e ao conjunto de dados limpos através do link (https://github.com/Melissa-Naldo/Crime\_In\_Portugal/blob/main/portugal\_crime\_2011-2019\_Organized.xlsx).

**Análise Exploratória de Dados:**

Este projeto de análise de dados concentra-se numa empresa de segurança que procura compreender os valores criminais em Portugal, especialmente crimes contra o património (incluindo assaltos a residências). Pode encontrar o ficheiro de consultas SQL em: https://github.com/Melissa-Naldo/Crime\_In\_Portugal/blob/main/SQL\_Crime.sql. Portanto, a primeira coisa que fiz foi analisar a minha tabela para compreender os meus dados (figura 7). Utilizando o seguinte código SQL:

SELECT

TABLE\_CATALOG,

TABLE\_SCHEMA,

TABLE\_NAME,

COLUMN\_NAME,

DATA\_TYPE

FROM INFORMATION\_SCHEMA.COLUMNS

O conjunto de dados consiste principalmente em colunas nvarchar, exceto "Year" e "Occurrences", que são do tipo float.

Figura 7 – As diferentes colunas e o seu tipo de dados.

**Análise geral dos dados:**

Para obter uma visão geral antes de me concentrar nos objetivos do projeto, foram realizadas várias consultas:

* Total de crimes em Portugal de 2011 a 2019:

Select Sum(Occurences) as Total\_Crimes, round(AVG(Occurences), 2) as Avg\_Crimes

From [dbo].[Crime Data]

* Quantidade de crime por região:

Select Region, SUM(Occurences) AS Sum\_Crimes\_per\_Region, round(AVG(Occurences), 2) AS AVG\_Crimes\_per\_Region, Rank() OVER (Order by SUM(Occurences) desc) as Rank

From [dbo].[Crime Data]

Group by Region

Order by AVG(Occurences) desc

* Quantidade de crimes por ano:

Select Year, Sum(Occurences) as Total\_Crimes\_Year, lag(Sum(Occurences)) Over (Order by Year) as Pre\_Year\_Crime, ((((Sum(Occurences))-(lag(Sum(Occurences)) OVER (Order by Year)))/(lag(Sum(Occurences)) OVER (Order by Year)))\*100) as Percentage\_Difference

From [dbo].[Crime Data]

Group by Year

Order by Year

Select Year, Sum(Occurences) as Total\_Crimes\_Year, lag(Sum(Occurences)) Over (Order by Year) as Pre\_Year\_Crime, ((((Sum(Occurences))-(lag(Sum(Occurences)) OVER (Order by Year)))/(lag(Sum(Occurences)) OVER (Order by Year)))\*100) as Percentage\_Difference

From [dbo].[Crime Data]

Where Year in (2011, 2016, 2017, 2019)

Group by Year

Order by Year

É fundamental analisar a evolução do crime total em várias regiões e ao longo dos anos. Após analisar os dados abrangentes desta investigação, optou-se por concentrar em regiões específicas, nomeadamente Norte, AM Lisboa, Algarve - regiões onde a empresa planeia abrir escritórios. Além disso, incluí o Centro, uma região que apresenta valores de crime notavelmente elevados. A consulta subsequente foi utilizada:

Select Region, Year, SUM(Occurences) AS Sum\_Crimes\_per\_Region, round(AVG(Occurences), 2) AS AVG\_Crimes

From [dbo].[Crime Data]

Where Region In ('Norte', 'AM Lisboa', 'Algarve', 'Centro')

Group by Region, Year

Order by Sum(Occurences) desc

**Crime contra o património:**

Após analisar todos os dados, decidiu-se dar um passo adicional e analisar os dados numa perspetiva diferente. Quis se observar valores semelhantes, mas aplicados apenas ao crime contra o património. Com isso em mente, foram feitas as consultas para encontrar o seguinte:

* Total de crimes contra o património:

Select Crime\_type, Sum(Occurences) as Total\_Patrimony\_Crimes, round(AVG(Occurences), 2) as Avg\_Crimes

From [dbo].[Crime Data]

Group by Crime\_type

* O total de crime contra o património de 2011 a 2019:

Select Year, Sum(Occurences) as Total\_Patrimony\_Crimes, (lag(Sum(Occurences)) OVER (Order by Year)) as Pre\_Year\_TotalCrime, ((((Sum(Occurences))-(lag(Sum(Occurences)) OVER (Order by Year)))/(lag(Sum(Occurences)) OVER (Order by Year)))\*100) as Percentage\_Difference

From [dbo].[Crime Data]

Where Crime\_type = ' Crimes against patrimony '

Group by Year

Order by Year

Select Year, Sum(Occurences) as Total\_Patrimony\_Crimes, (lag(Sum(Occurences)) OVER (Order by Year)) as Pre\_Year\_TotalCrime, ((((Sum(Occurences))-(lag(Sum(Occurences)) OVER (Order by Year)))/(lag(Sum(Occurences)) OVER (Order by Year)))\*100) as Percentage\_Difference

From [dbo].[Crime Data]

Where Crime\_type = ' Crimes against patrimony ' and Year in (2011, 2019)

Group by Year

Order by Year

* As quatro principais regiões com mais crimes contra o património:

Select Region, Sum(Occurences) as Total\_Patrimony\_Crimes, round(AVG(Occurences), 2) as Avg\_Patrimony\_Crimes

From [dbo].[Crime Data]

Where Crime\_type = ' Crimes against patrimony '

Group by Region

Order by SUM(Occurences) desc

* Evolução dos crimes contra o património nas quatro principais regiões com mais crimes ao longo de 9 anos:

Select Region, Year, SUM(Occurences) AS Sum\_Patrimony\_Crimes, round(AVG(Occurences), 2) AS AVG\_Crimes\_patrimony

From [dbo].[Crime Data]

Where Crime\_type = ' Crimes against patrimony '

Group by Region, Year

Having Region In ('Norte', 'AM Lisboa', 'Algarve', 'Centro')

Order by Sum(Occurences) desc

* Principais municípios com mais crime em 2019:

Select Year,

Region,

Municipality,

Occurences

From [dbo].[Crime Data]

Where Crime\_type = ' Crimes against patrimony 'AND Year=2019

Order by Occurences desc

**Correção de dados:**

No entanto, ao observar corretamento os dados, vejo que os valores médios para o total de crimes e para os tipos de crimes são extremamente baixos. Isso acontece porque o SQL está a dividir o meu valor total com todos os meus valores das diferentes localizações. Portanto, é necessário encontrar meu valor médio. Utilizei as seguintes consultas para fazer isso:

----Total

--Correcting the overall average

With Total\_crime\_avg as(Select

Year,

Sum(Occurences) as total

From [dbo].[Crime Data]

Group by Year

)

Select round(avg(total), 2) as avg\_crimes\_per\_year

From Total\_crime\_avg

-- Correcting the average per Region

Select Region, Count(Distinct(Municipality)) as Nº\_municipalities, SUM(Occurences) AS Sum\_Crimes, round((SUM(Occurences)/Count(Distinct(Municipality))), 2) AS AVG\_Crimes

From [dbo].[Crime Data]

Group by Region

Order by round((SUM(Occurences)/Count(Distinct(Municipality))), 2) desc

-----Patrimony - I'm going to do the same correction but focused on my patrimony crime

--Correcting the overall patrimony crime average

With Patrimony\_crime\_avg as (Select

Year,

Sum(Occurences) as total

From [dbo].[Crime Data]

Where Crime\_type = ' Crimes against patrimony '

Group by Year

)

Select round(avg(total), 2) as avg\_patrimony\_crimes

From Patrimony\_crime\_avg

--Correcting my overall average per region

Select Region, Count(Distinct(Municipality)) as Nº\_municipalities, SUM(Occurences) AS Sum\_Crimes, round((SUM(Occurences)/Count(Distinct(Municipality))), 2) AS AVG\_Crimes

From [dbo].[Crime Data]

Where Crime\_type = ' Crimes against patrimony '

Group by Region

Order by round((SUM(Occurences)/Count(Distinct(Municipality))), 2) desc

-- Correcting the average per region and year

Select Year, Region, Count(Distinct(Municipality)) as Nº\_municipalities, SUM(Occurences) AS Sum\_Patrimony\_Crimes, round((SUM(Occurences)/Count(Distinct(Municipality))), 2) AS AVG\_Patrimony\_Crimes

From [dbo].[Crime Data]

Where Crime\_type = ' Crimes against patrimony '

Group by Region, Year

Having Region In ('Norte', 'AM Lisboa', 'Algarve', 'Centro')

Order by round((SUM(Occurences)/Count(Distinct(Municipality))), 2) desc

Após encontrar os valores corretos, criou-se tabelas para armazenar dados importantes para que,, posteriormente, fosse possível aceder na fase de visualização do projeto. Criou-se a tabela "Updated\_Crime\_Average" com a média corrigida total dos crimes e a média dos crimes contra o património. Utilizei a seguinte consulta para alcançar esse objetivo:

Create Table Updated\_Crime\_Average (

Avg\_Total decimal(8, 2),

Avg\_Patrimony decimal(8, 2))

Insert into Updated\_Crime\_Average Values

(348254.72, 189197.11)

Select \*

From Updated\_Crime\_Average

Criou-se uma tabela com a média correta de crimes para as diferentes regiões, mas criou-se esta manualmente no Object Explorer e editou-se as linhas para inserir os valores corretos (figura 8).

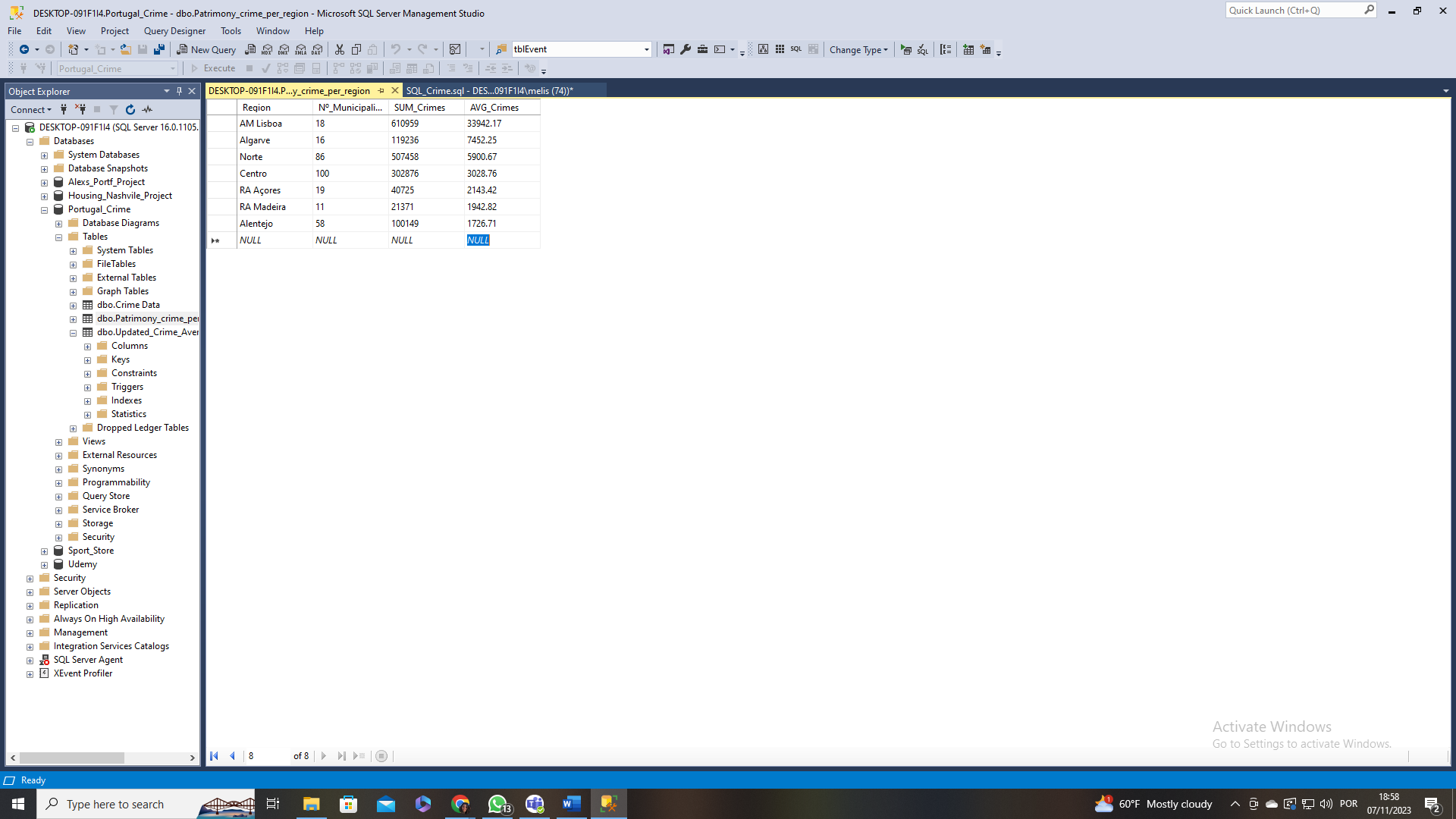


Figura 8 - Tabela com valores médios atualizados para as diferentes regiões.

**Criação das viasualizações**

Para criar as visualizações, optou-se por organizar os resultados em duas páginas distintas no painel Power BI. A primeira página visa oferecer uma compreensão abrangente da evolução da criminalidade de 2011 a 2019, enquanto a segunda página aborda diretamente o objetivo principal do projeto.

Os dados foram importados do SQL para o Power BI, incorporando as três principais tabelas, juntamente com uma tabela adicional contendo as coordenadas para diferentes regiões em Portugal. Garantir relacionamentos corretos entre essas tabelas, usando a coluna "Region" como uma chave de ligação, foi crucial.

***Visualizações:***

As visualizações foram criadas usando uma variedade de elementos, tais como:

* Cartões: Utilizados para mostrar informações cruciais.
* Gráficos de linhas: Empregados para ilustrar a evolução dos dados ao longo dos anos.
* Gráficos de barras: Utilizados para destacar regiões com as taxas de crime mais elevadas.
* Mapas: Utilizados para identificar locais específicos.

***Dashboard:***

O relatório final, elaborado no Power BI, pode ser acedido através do link: https://github.com/Melissa-Naldo/Crime\_In\_Portugal/blob/main/Crime\_Report.pbix

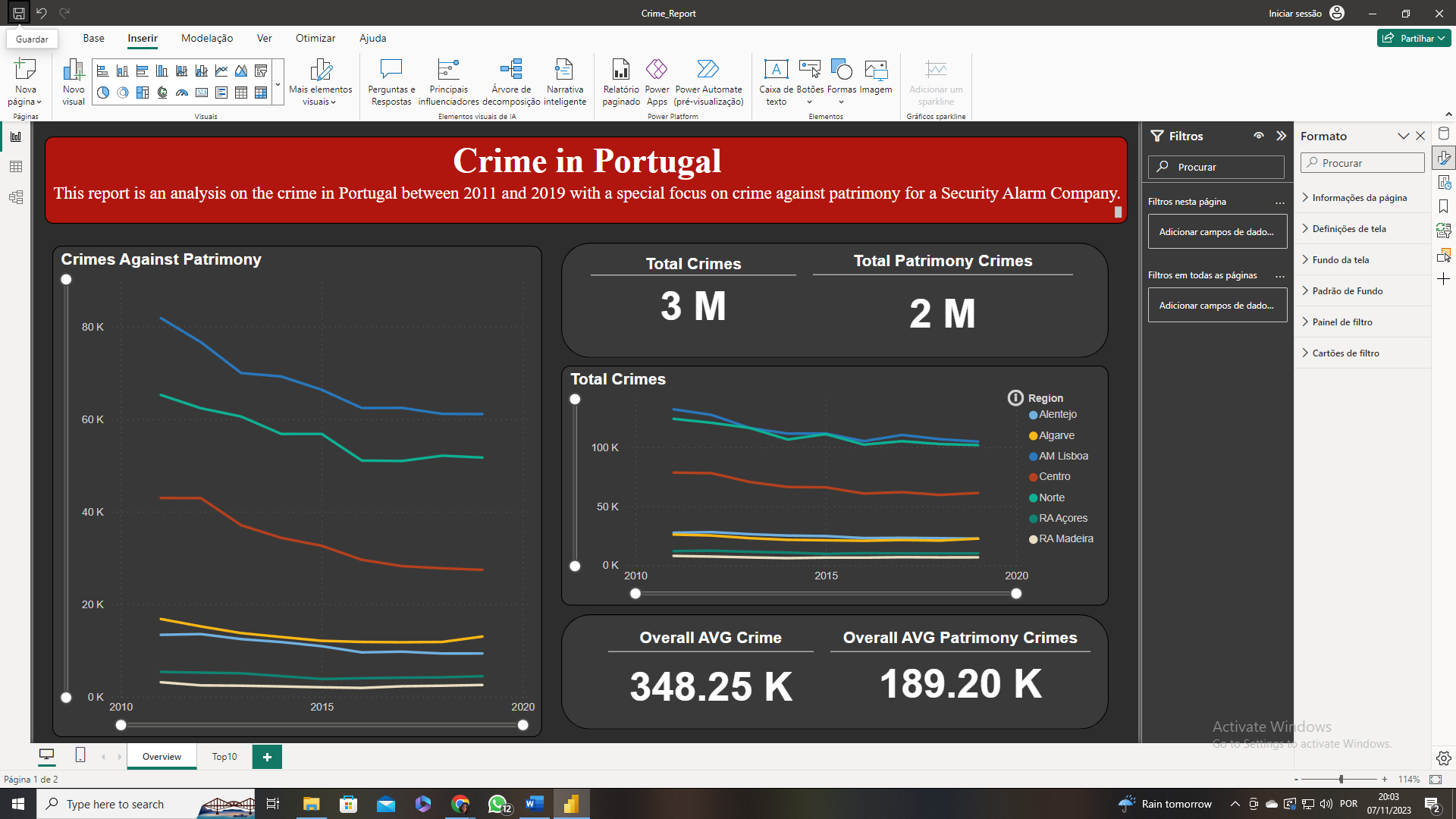


Figura 9 – Primeira página do dashboard do crime em Portugal.

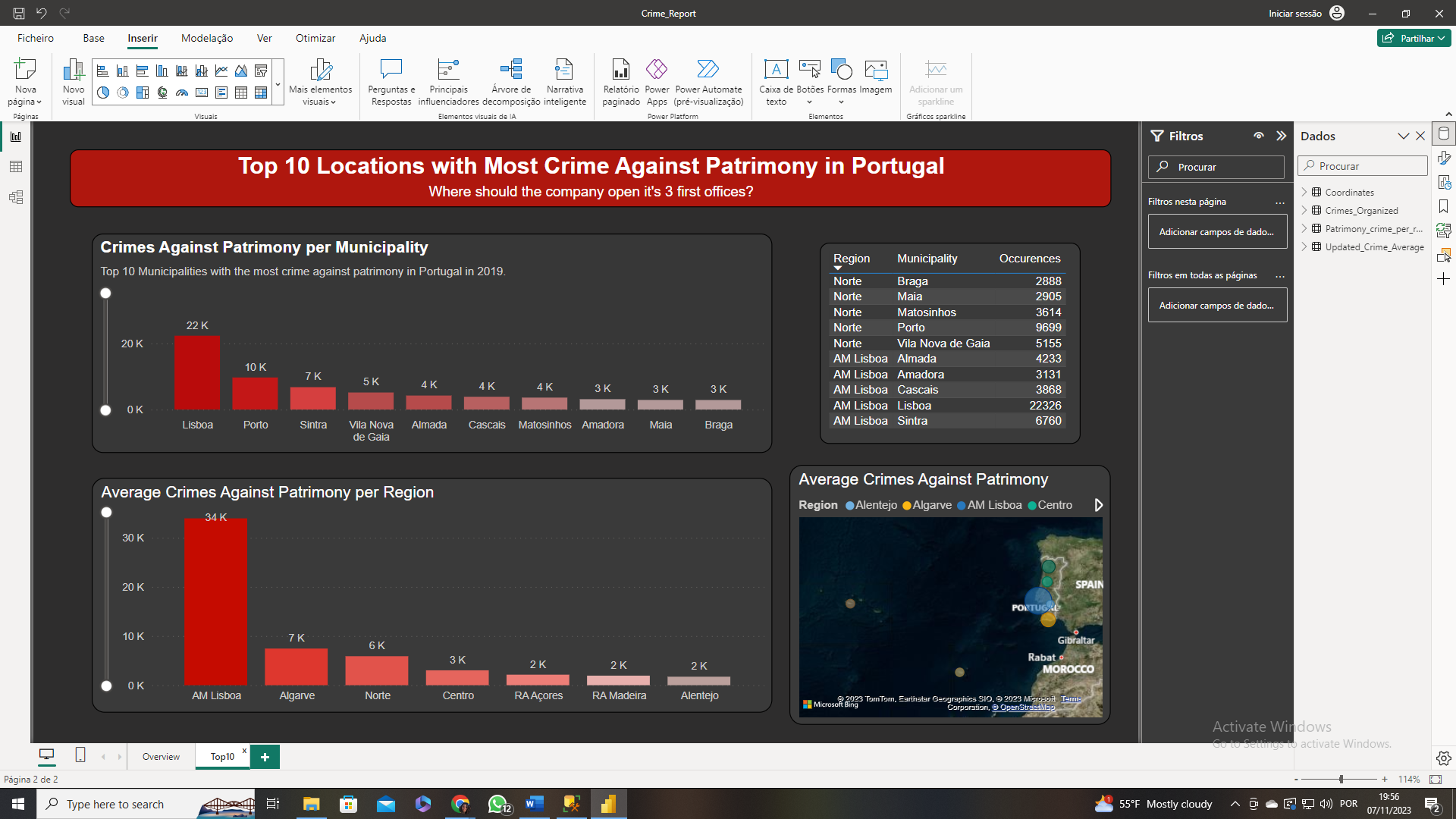


Figura 10 – Segunda página do dashboard do crime em Portugal.

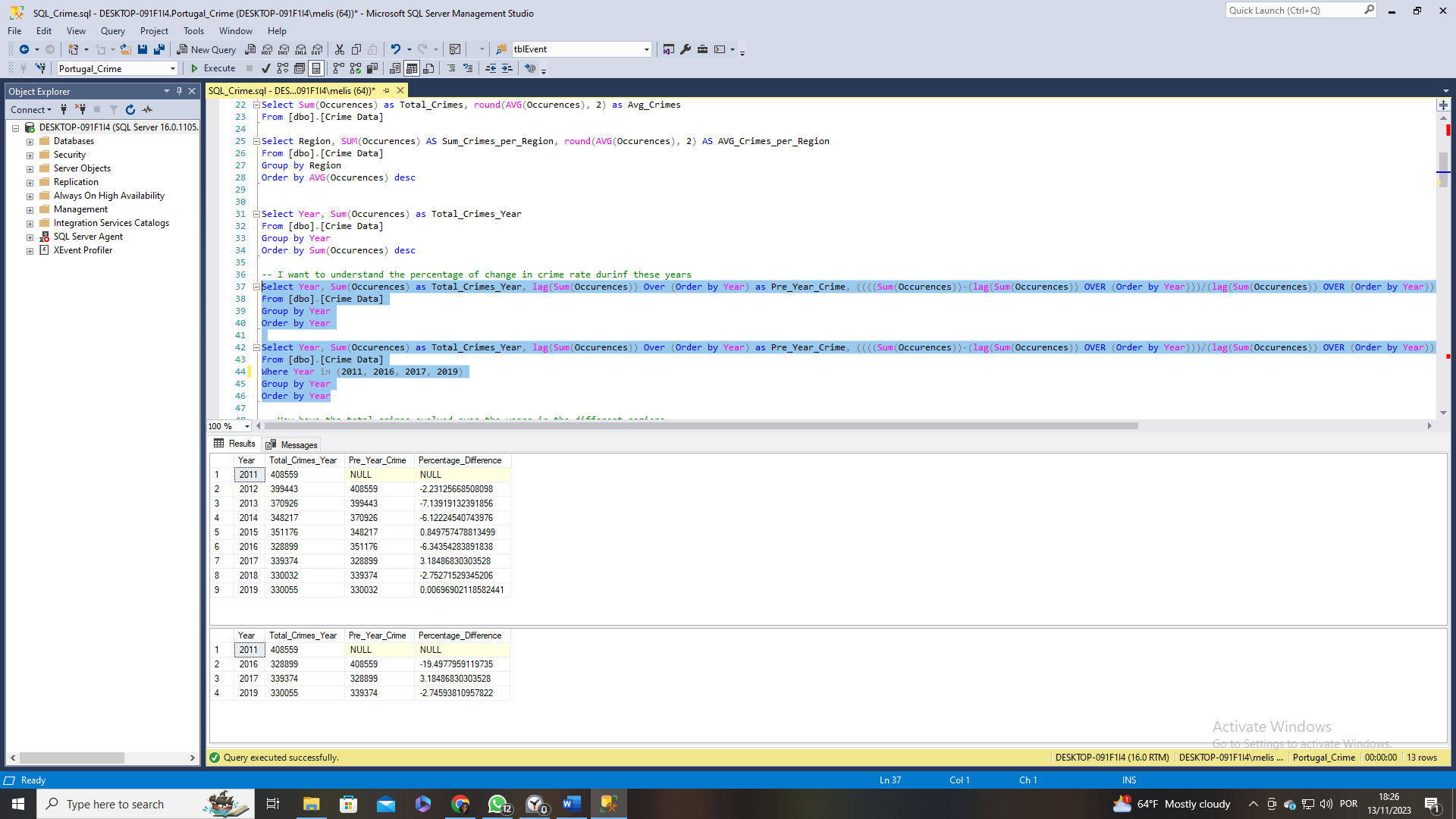
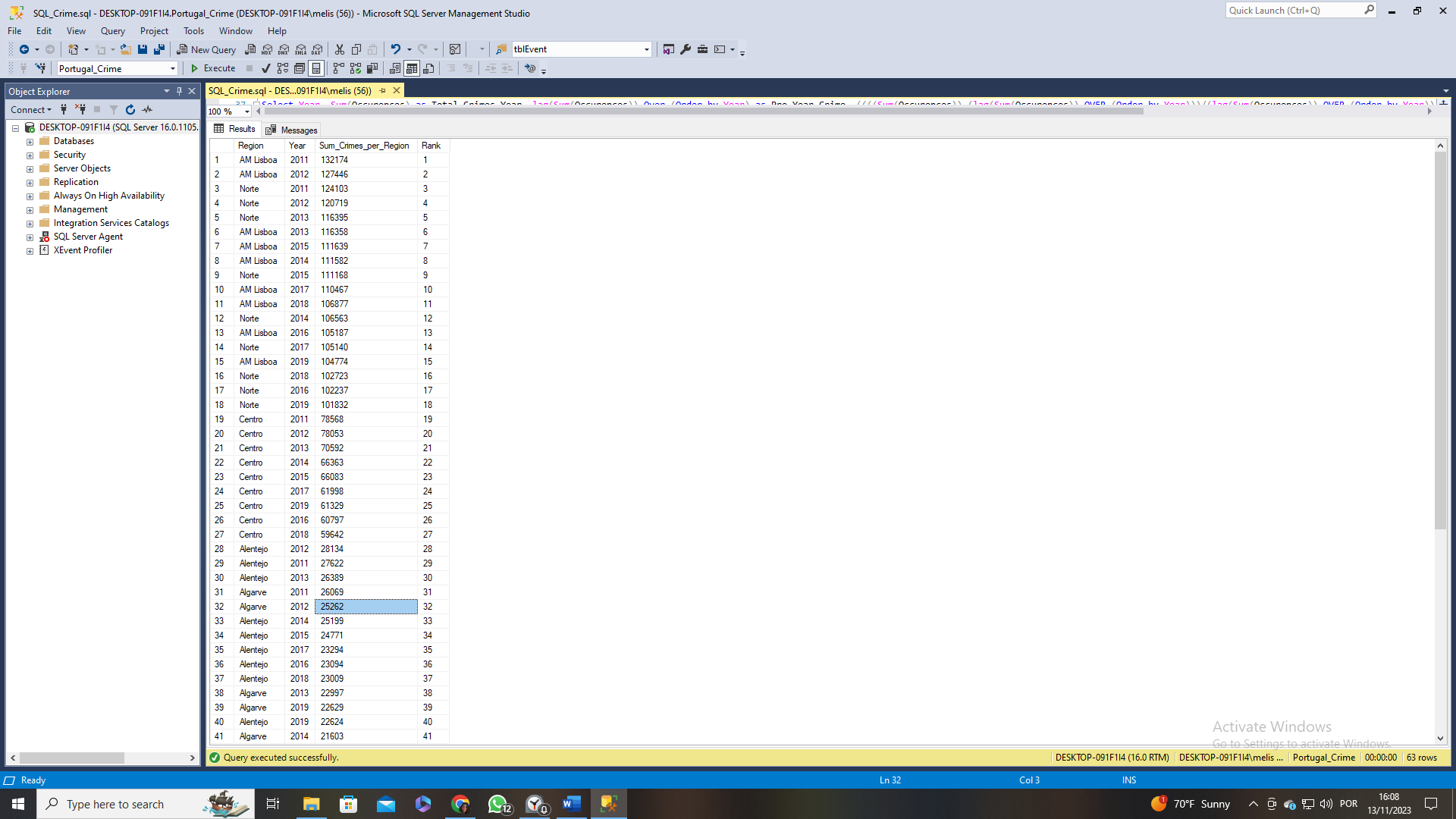
Resultados

Figura 11 – Quantidade de crime por ano e a sua evolução.

Após analisar o conjunto de dados limpo, pude obter resultados significativos. De uma forma geral, o número total de crimes relatados em Portugal entre 2011 e 2019 foi 3 206 681, com uma média anual de 356 297,89.

No entanto, adicionando a componente temporal aos valores, observou-se que o crime total em Portugal começou com um valor elevado (408 559) e começou a diminuir até atingir 328 899 em 2016 (uma redução de 19,49%), antes de aumentar novamente, atingindo um pico em 2017 com 339 374 crimes totais (um aumento de 3,18%). Após 2017, a taxa de criminalidade começou novamente a declinar até 2019. Curiosamente, após 2015, os valores totais de crimes permaneceram consistentemente abaixo da média. A imagem à direita mostra a quantidade total de crimes por ano e sua evolução.

 Portugal possui 7 regiões, juntas elas representam os valores mencionados anteriormente, mas os 4 locais com a maior média de crime são “AM Lisboa”, “Algarve”, “Norte” e “Centro” (representados na Figura 13). No entanto, foi dado um passo adicional nesta análise ao consultar o ranking do valor total de crime por ano e por região. Isso permitiu compreender que “AM Lisboa” é a região com mais crime, tendo em 2011 registado o valor mais alto e em 2012 o seu segundo mais alto; em seguida, em 3º lugar temos o “Norte” com um total de 124 103 em 2011; o “Centro” vem em 19º com 78 568 em 2011; e “Algarve” aparece apenas em 32º lugar com 25 262. Podemos ver os primeiros 20 lugares na Figura 12.

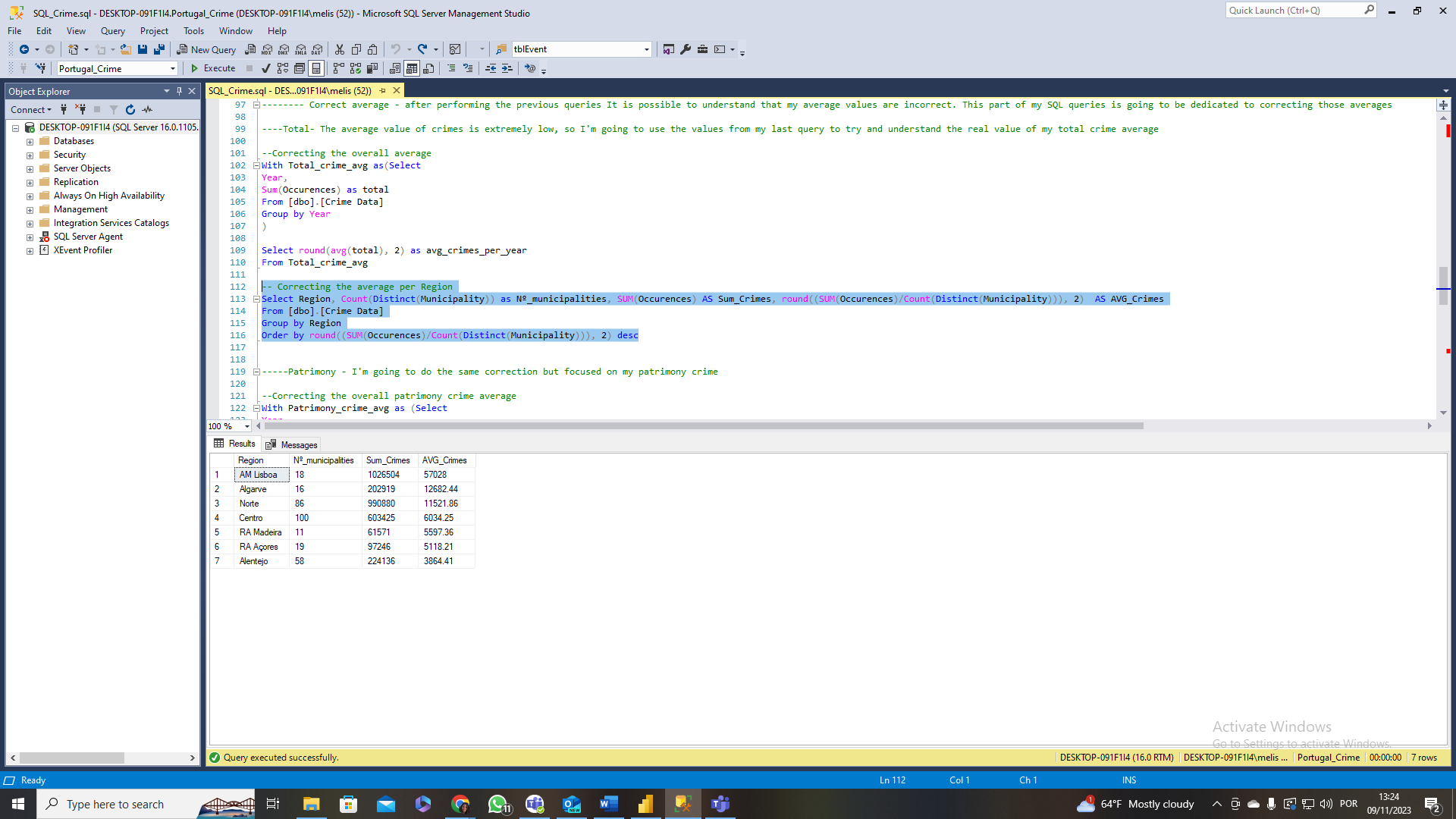


Figura 12 – Ranking do total de crimes por região e por ano.

Figura 13 – Quantidade de crime por região.

Após obter uma visão geral dos dados, analisou-se o tipo de crimes e viu-se a quantidade de crime contra o património existente em Portugal (Figura 12). Olhando para os dados, representados na Figura 14, é possível observar que o crime contra o património é o mais cometido em Portugal. Notavelmente, mais da metade dos crimes totais se enquadram na categoria de crimes contra o património, alcançando um valor de 1 702 774, com uma média de 189 197,11 crimes por ano.

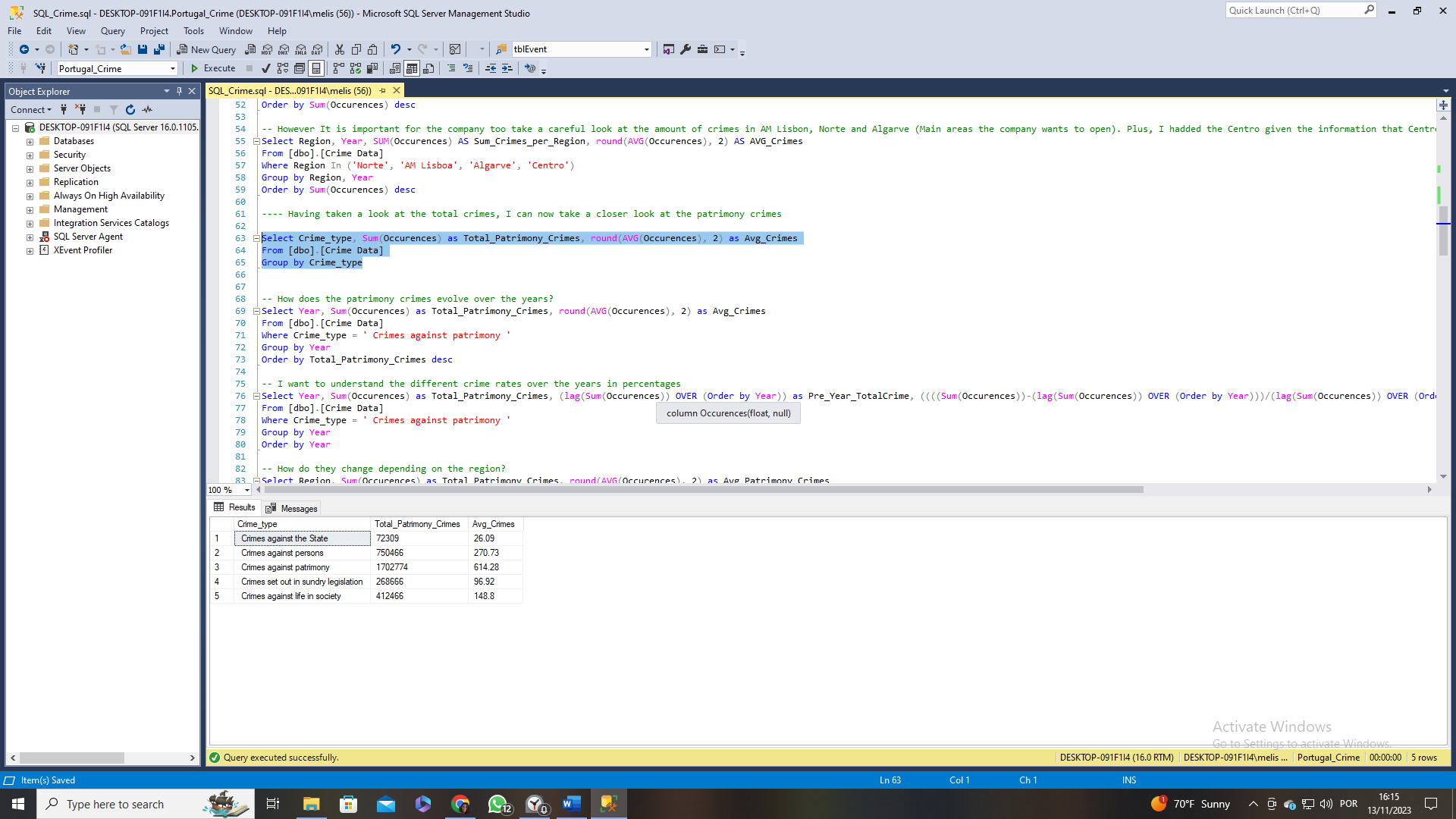


Figura 14 – Tipos de crime em Portugal.

No entanto, o crime contra o património tem diminuído constantemente ao longo dos anos (Figura 15), começando com um valor de 228 704 em 2011 e terminando com um valor de 169 548 em 2019, uma redução de 25,86%. Notavelmente, é possível ver que os crimes contra o património de 2016 a 2019 estão abaixo da média.

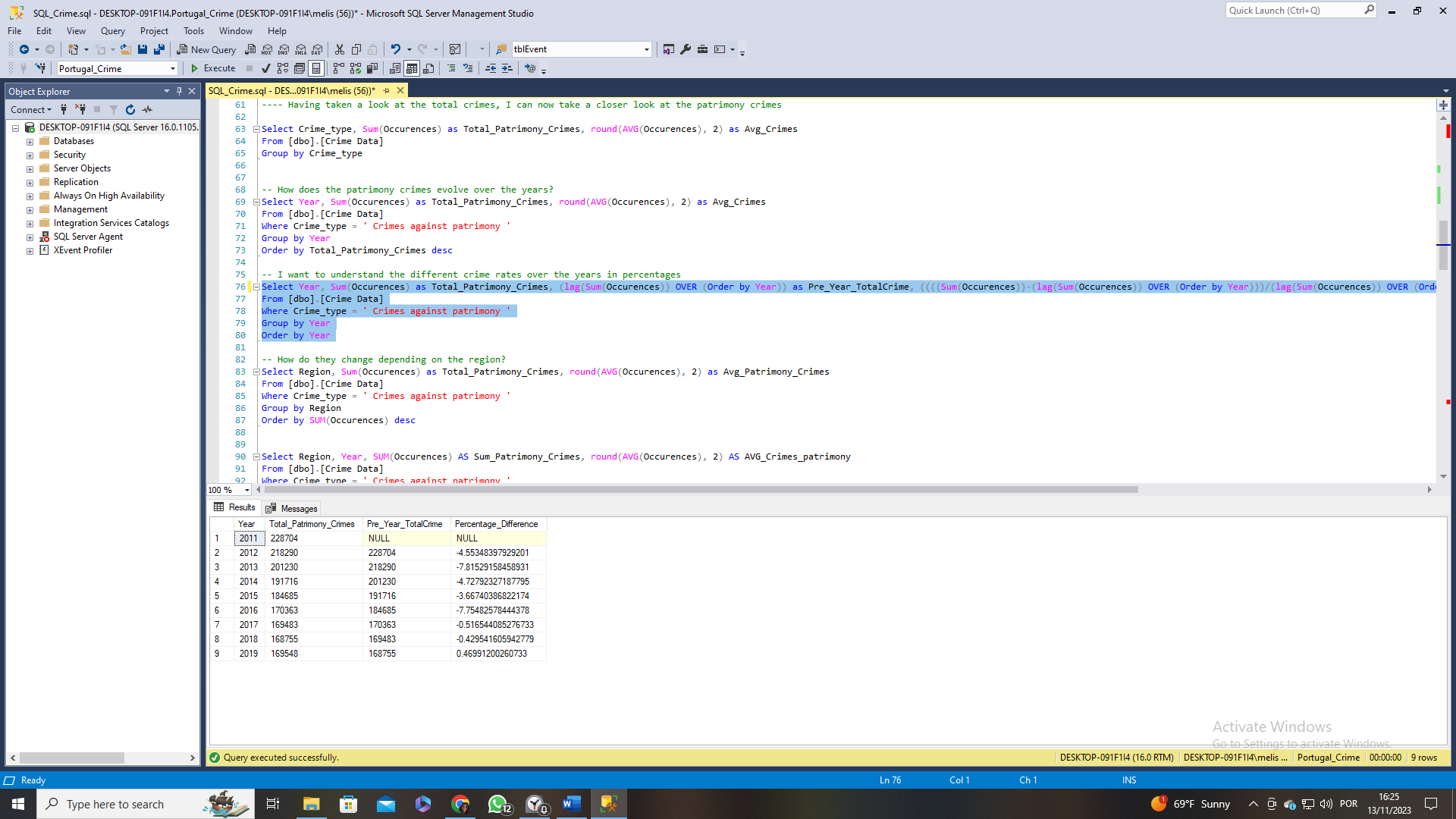


Figura 15 - Valor total de crimes contra o património entre 2011 e 2019 e a sua evolução percentual.

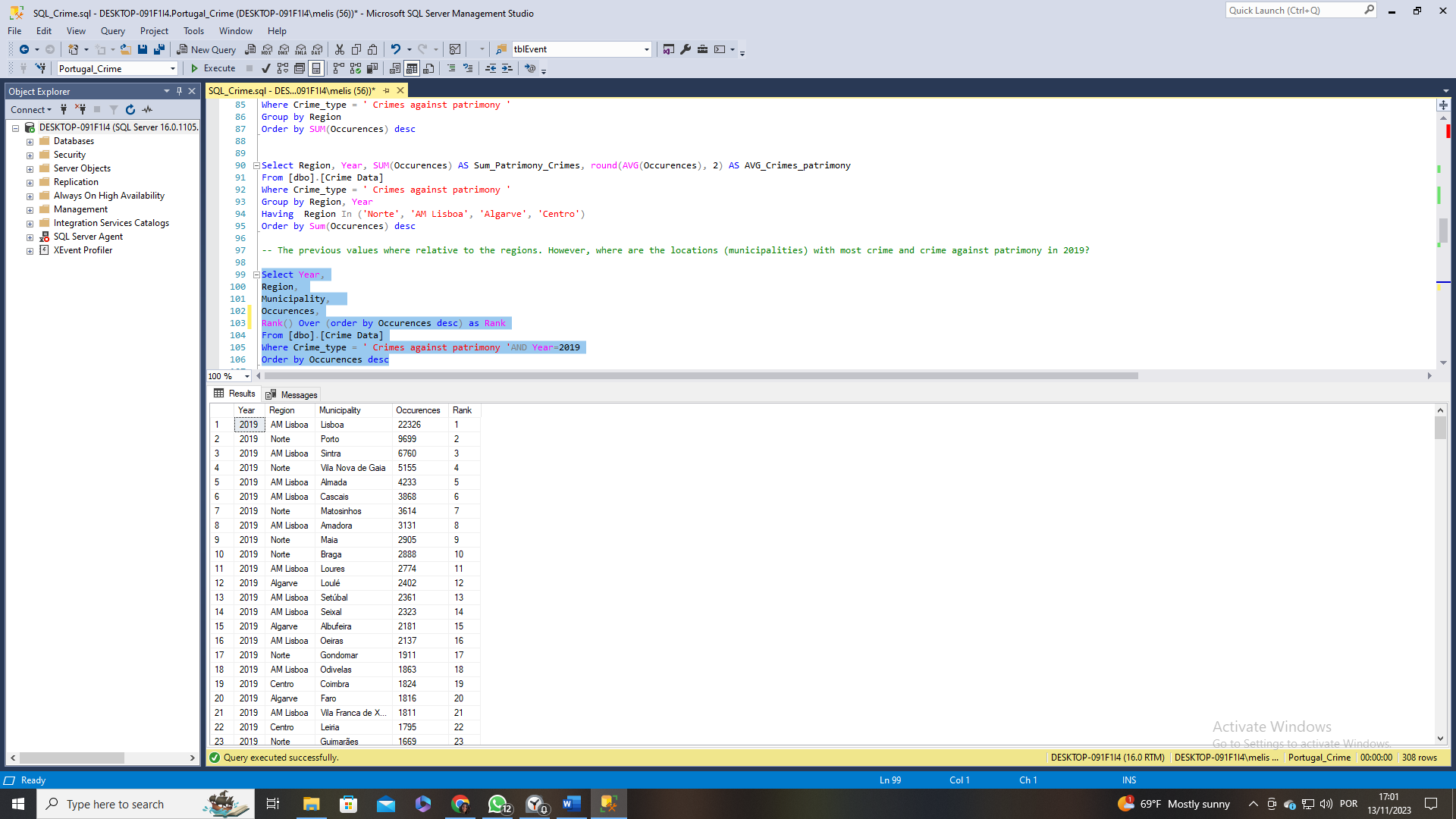
Ao analisar mais especificamente quais as regiões e municípios que possuem mais crime contra o património, obteu-se as informações representadas na Figura 16 e 17. As 4 principais regiões com mais crime contra o património são “AM Lisboa”, "Algarve”, “Norte” e “Centro”. No entanto, olhando para locais específicos, usando a coluna de município, é interessante notar que em 2019 os municípios com as taxas de crime mais altas pertencem a “AM Lisboa”, depois “Norte”, “Algarve” e depois “Centro”.

Figura 16 - Os 20 top municípios com mais crime contra o património em 2019.

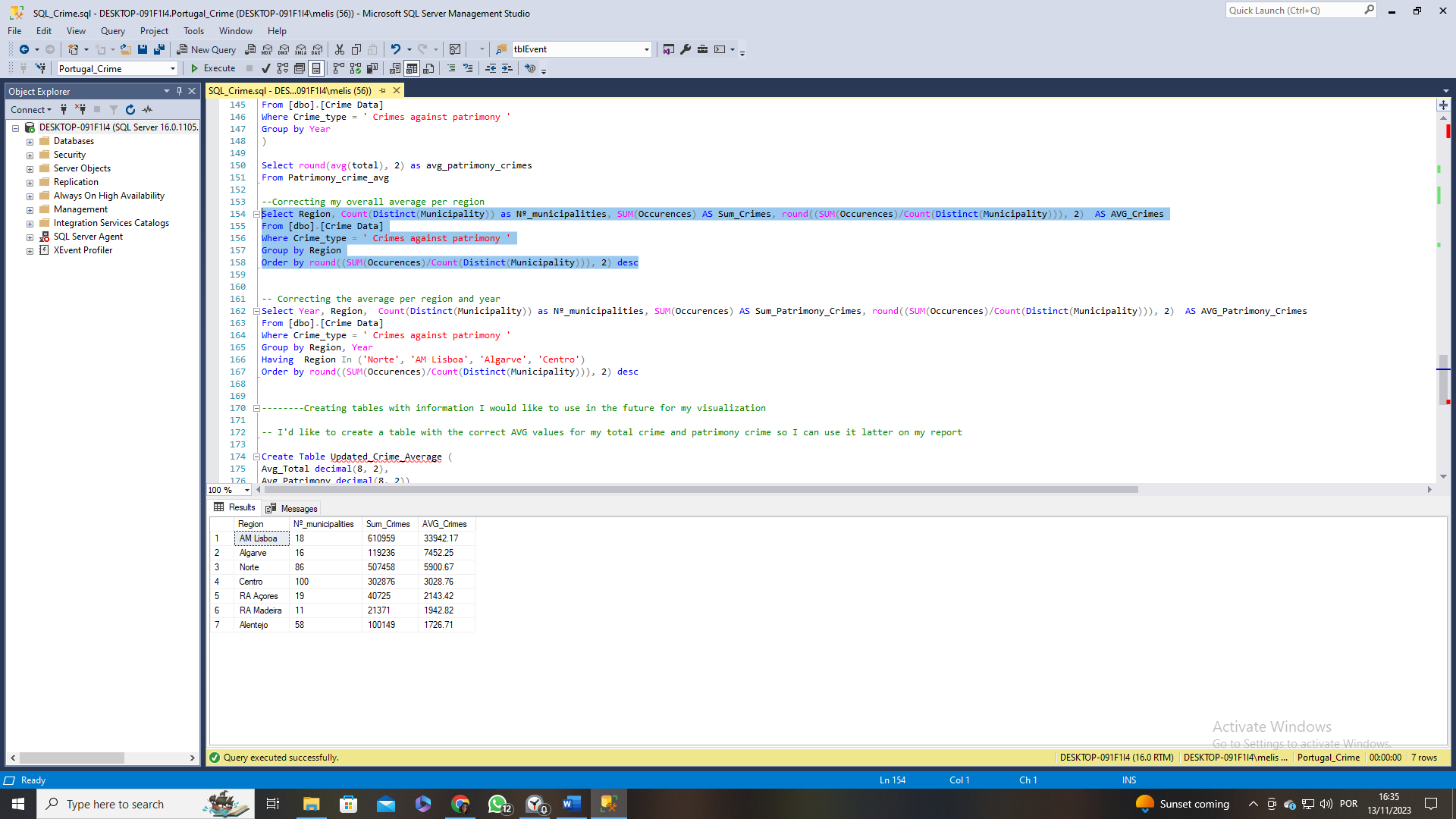


Figura 17 - Valores de crime contra o património por região Portuguesa.

Conclusão

O objetivo deste projeto foi determinar as regiões ideais em Portugal para uma empresa de segurança residencial estabelecer os seus primeiros escritórios. Com foco em três cidades principais—Lisboa (AM Lisboa), Porto (Norte) e Algarve—a análise procurou compreender a situação geral de criminalidade em Portugal e, mais especificamente, crimes contra o património.

Após o processo de análise, pudemos observar que entre 2011 e 2019 foram relatados 3 206 681 crimes, com uma média de 356 298 por ano. De todos os crimes, mais da metade foram crimes contra o património, atingindo 1 702 774, e com uma média de 189 197 crimes por ano. A criminalidade geral teve uma diminuição de 19,5% de 2011 a 2016; tendo nos anos seguintes, ocorrido um aumento de 0,35% na criminalidade total. No entanto, os crimes contra o património tiveram uma redução significativa de 25,86%, apesar do aumento da criminalidade total de 2016 a 2019.

Focando ainda mais nas diferentes regiões, sabemos que aquelas com mais crimes e crimes contra o património são “AM Lisboa”, “Norte”, “Algarve” e “Centro”. A Figura 18 representa os valores de criminalidade dessas diferentes regiões:

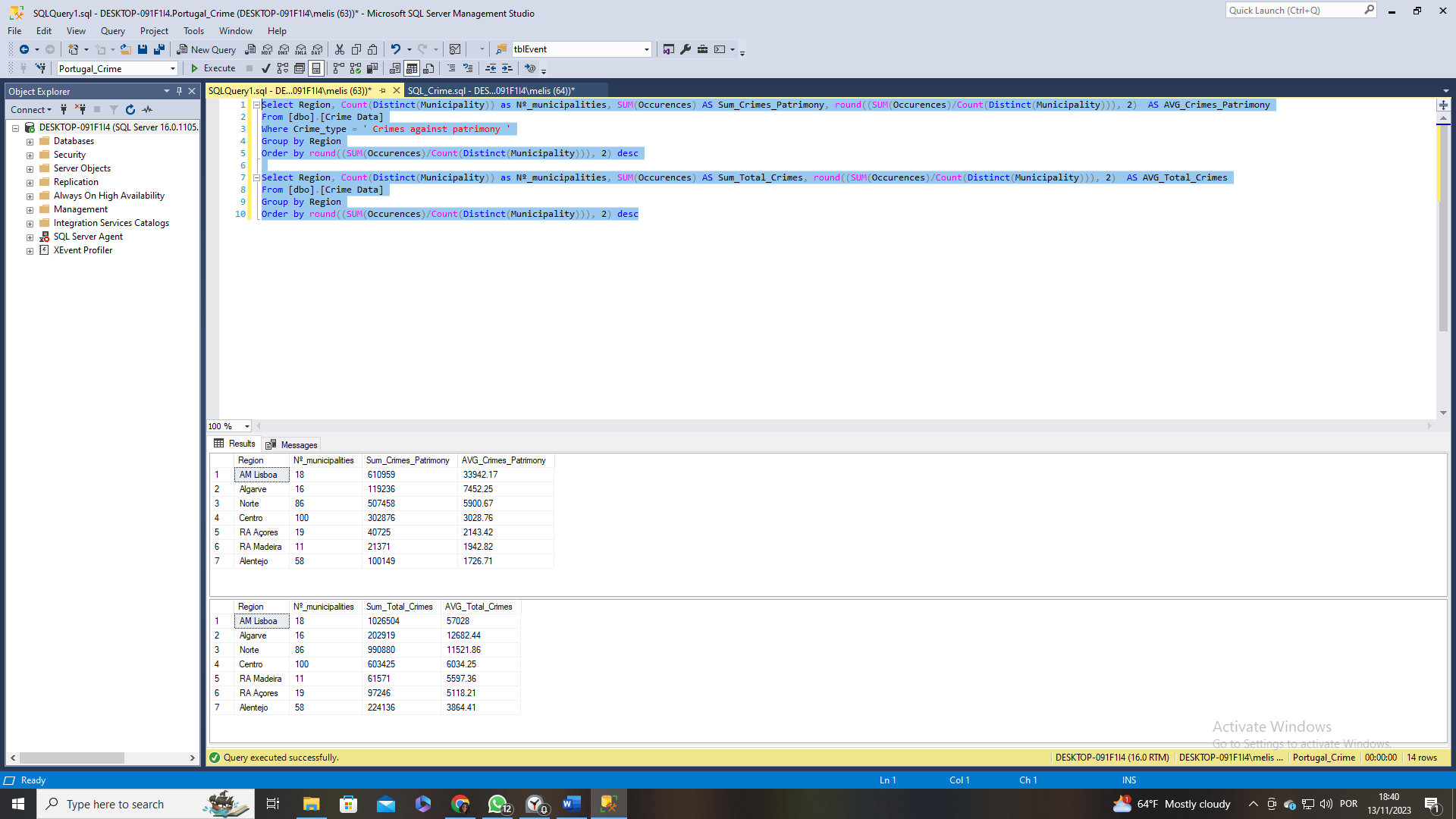


Figura 18 - Comparação de crime total e crimes contra o patrimônio entre as quatro principais regiões.

Se concentrarmos nos valores médios, podemos classificar Lisboa como a principal localização para abrir os escritórios, seguida pelo Algarve e Norte. Isso excluiria a região Centro, que possui um valor total de crimes significativamente maior do que o Algarve, apesar de sabermos que o Algarve é significativamente menor, com apenas 16 municípios em comparação com os 100 municípios do Centro.

Indo ainda mais a fundo nesta análise, ao observar de perto os municípios, sabemos que os 10 principais municípios com mais crimes pertencem a AM Lisboa e Porto. Quanto à região do Algarve, o município com mais crimes está classificado em 12º lugar na lista de municípios com mais crimes contra o património.

Com base na análise, AM Lisboa, seguida por Norte e Algarve, emerge como as localizações mais favoráveis para abrir escritórios. Essas localizações seriam mais benéficas porque apresentam os maiores valores médios de crime. Embora a análise aponte para regiões específicas, é crucial adaptar as recomendações com base em dados em tempo real, políticas locais e na natureza em evolução da criminalidade. Uma compreensão abrangente da dinâmica populacional melhoraria ainda mais a avaliação.