

**Business Documentation**

**São Paulo, 26 de janeiro de 2021.**

|  |  |  |  |
| --- | --- | --- | --- |
| **CONTROLE DE VERSÃO** | | | |
| **Autor** | **Versão** | **Data** | **Descrição** |
| **Elizangela S. Pinheiro**  **Flavia B. Pereira**  **Igor Rodrigues**  **Isabela Peron** | 1.0 | 26/01/2021 | Criação do documento. |
| **Danilo Donato**  **Diego Botelho de Souza** | 1.1 | 28/01/2021 | Ajustes nas informações |
| **Isabela Peron** | 1.2 |  | Finalização do documento. |

1. **Introdução**

O seguinte documento tem por objetivo detalhar os requisitos técnicos seguidos para a elaboração do Projeto Integrado proposto pelo programa Blueshift Academy como forma de aprendizado.

1. **Solicitação**

O projeto visa identificar e apresentar informações voltadas aos boletins eletrônicos de ocorrências de feminicídio no estado de São Paulo entre o período de Janeiro/2016 e Dezembro/2020.

O desenvolvimento do projeto ocorre a partir da obtenção dos dados através do portal da **Secretaria de Estado da Segurança Pública** utilizando-se Python + Selenium para a automação do processo de download dos arquivos e armazenamento dos cujo ditos no Google Cloud Plataform(GCP) Storage.

Já na segunda parte do projeto, realiza-se a modelagem dos dados vindo do Storage no BigQuery e a criação do dashboard no Data Studio.

1. **Premissas da Solução** 
   1. **Origem e especificação dos dados**

Os arquivos utilizados no desenrolar do projeto têm como origem o portal da Secretária da Segurança Pública de SP baixados de forma automatizada no formato .xls, totalizando 60itens, seguindo com a nomenclatura:

DadosBO\_Ano\_Mês(FEMINICÍDIO).xls

Sendo “Ano” o ano de referência do arquivo, que varia de 2016 à 2020 e “Mês” ao mês de referência, podendo este transitar de 01 à 12

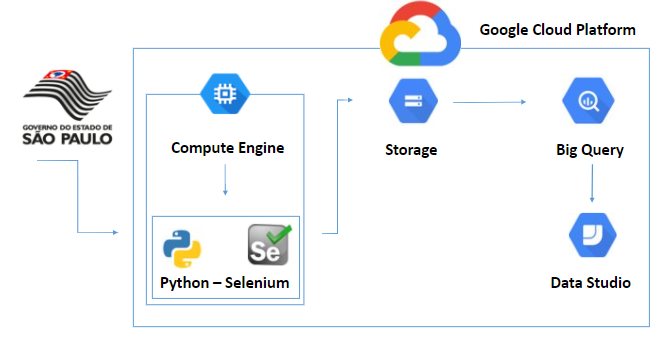
* 1. **Ambiente de Desenvolvimento**

Para a elaboração do projeto, foi escolhido a plataforma Google Cloud Platform, onde, para tanto, a empresa Blueshft disponibilizou os acessos às ferramentas necessárias para o desenvolvimento com as ferramentas abaixo proposta:

* **Computer Engine –** Criação da VM Windows e elaboração do script em Python + Selenium
* **Storage –** Armazenamento do arquivo a ser analisado
* **BigQuery –** Modelagem e tratamento dos dados
* **Data Studio –** Elaboração dos Dashboards.

1. **Modelo da Arquitetura Sugerida**

Abaixo, uma representação da arquitetura escolhida para a solução de acordo com o entendimento e levantamento dos requisitos



**1**

**5**

**3**

**4**

**2**

1. **Portal Segurança Publica**

<http://www.ssp.sp.gov.br/transparenciassp/Consulta.aspx>

1. **Computer Engine** 
   1. **Criação da VM Windows**

|  |  |
| --- | --- |
| **Propriedades** | **Valor** |
| Nome | Proj-integrado-vm |
| Série | N1 |
| Tipo de Máquina | N1-standard-8(vCPU, 30 GB de memória) |
| Disco de Utilização | Windows Server 2019 Datacenter |
| Tamanho do Disco | 50 GB |
| Identidade e Acesso a API | Permitir acesso padrão |
| Firewall | Permitir trafégo HTTP e HTTPS |

* 1. **- Softwares Utilizados**
  + Google Chrome - Navegador
  + Jupyter Notebook – Ambiente de desenvolvimento de scripts Python

(<https://github.com/ElizangelaS/Projeto-Integrado/tree/master/ScriptPython>)

* + Chrome WebDriver – Plugin de automatização do navegador Google Chrome
  + Microsoft Excel – Gerenciador de planilhas

1. **Cloud Storage** – Serviço para o armazenamento do arquivo
   1. Propriedades do Bucket criado para o projeto

|  |  |
| --- | --- |
| **Propriedades** | **Valor** |
| Nome Bucket | Proj\_integrado\_stg |
| Tipo de Local | Region |
| Local | us-east1 (Carolina do Sul) |
| Classe de Armazenamento Padrão | Standard |
| Controle de Acesso | Detalhado |
| Criptografia | Chave gerenciada pelo Google |

1. **BigQuery** – Manipulação e tratamento dos dados
   1. Criação conjunto de dados:

|  |  |
| --- | --- |
| Propriedades | Valor |
| Id Conjunto de dados | BaseConsolidada |
| Local dos Dados | Padrão |
| Expiração da Tabela Padrão | Nunca |
| Criptografia | Chave gerenciada pelo Google |

* 1. Criação da Tabela:

Origem:

|  |  |
| --- | --- |
| Criar tabela de: | Google Cloud Storage |
| Bucket do GCS | gs://proj\_integrado-stg |
| Formato do Arquivo | CSV |

Destino:

|  |  |
| --- | --- |
| Nome projeto | blueshift-academy |
| Nome conjunto de dados | BaseConsolidada |
| Tipo de tabela | Tabela Nativa |
| Nome da tabela | BaseFem |
| Esquema | Detectar automaticamente |

* 1. Procedure para tratamento de campos nulos e Script de criação e inserção na tabela resumida para o consumo no Data Studio

<https://github.com/ElizangelaS/Projeto-Integrado/tree/master/ScriptSQL>

1. **Modelagem**

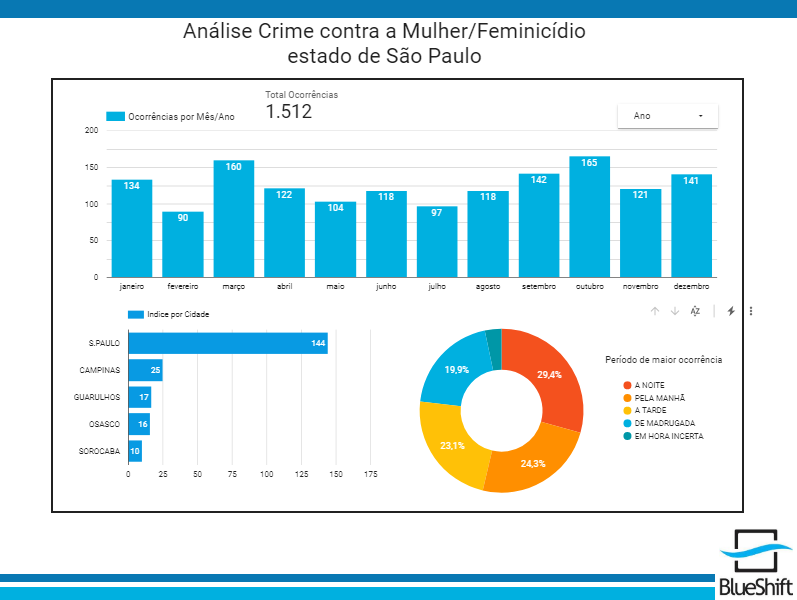
O modelo abaixo representa o schema da tabela final utilizada:

|  |  |
| --- | --- |
| **NUM\_BO** | **INTEGER** |
| **DATAOCORRENCIA** | **DATE** |
| **PERIDOOCORRENCIA** | **STRING** |
| **BO\_AUTORIA** | **STRING** |
| **CIDADE** | **STRING** |
| **UF** | **STRING** |
| **DESCRICAOLOCAL** | **STRING** |
| **RUBRICA** | **STRING** |
| **STATUS** | **STRING** |
| **NOMEPESSOA** | **STRING** |
| **ESTADOCIVIL** | **STRING** |
| **GRAUINSTRUCAO** | **STRING** |
| **CORCUTIS** | **STRING** |
| **NATUREZAVINCULADA** | **STRING** |

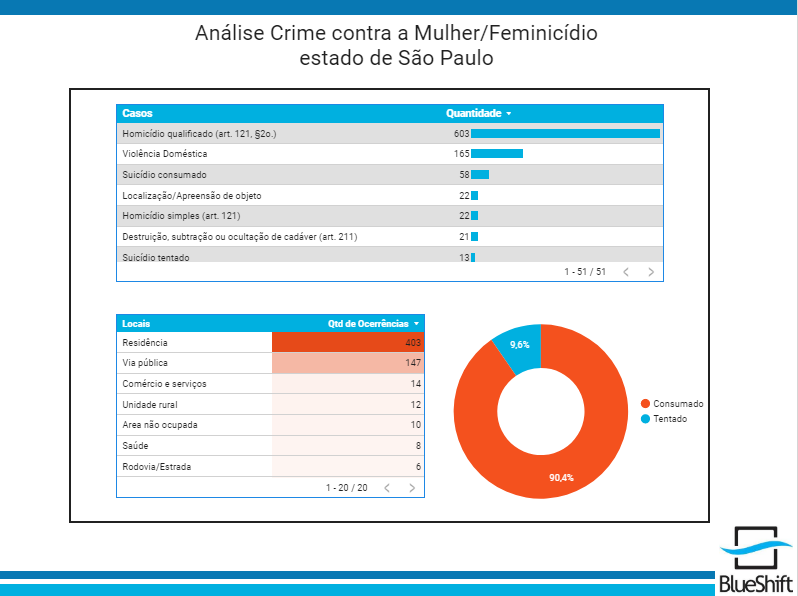
1. **Dashboard**

Os dados que foram tratados e manipulados no Big Query foram enviados para o Data Studio, em que foram realizadas as dashboards/view como demonstrados abaixo:

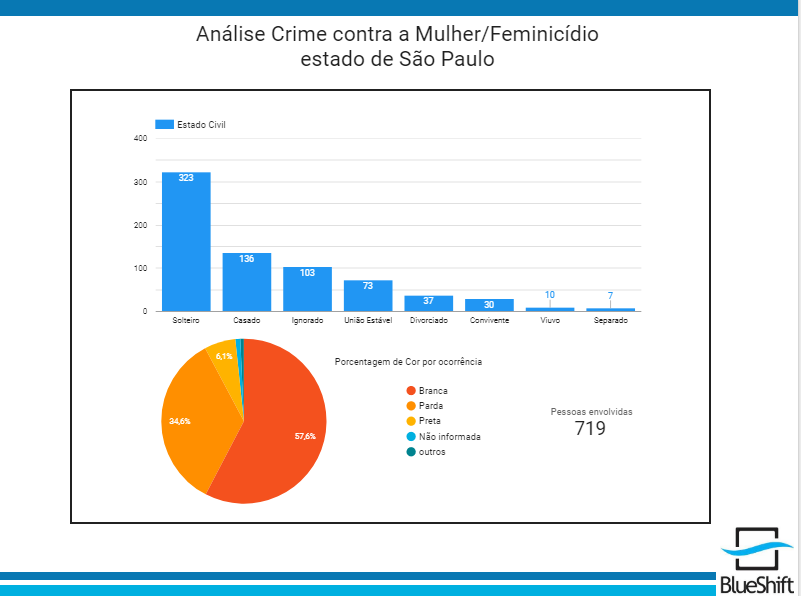
Página 1:



Página 2:



Página 3:

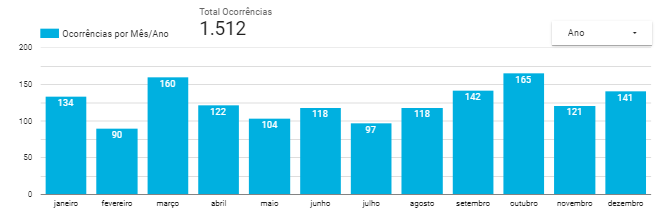


* **Ocorrência de casos de feminicídio por ano/meses:**

Baseado nas informações das bases de dados, foi possível obter os seguintes dados:

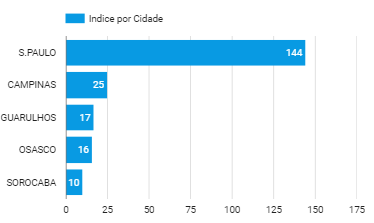
* *Cards* para demonstração do Total de Registros (métrica automática).
* Gráfico com representação dos números de boletins que foram realizados por ano. Sendo utilizado o controle para escolha do ano (2016-2020).
* Comparativo mensal do número de boletins de ocorrência.

Figura 1: Gráfico representando o número de boletins por mês/ano. E visão geral do total de registros, pessoas envolvidas e ocorrências distintas.



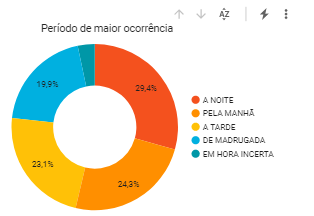
* **Boletins de ocorrência por cidades:**
* Gráfico de barra para representação do número de boletins de ocorrências registrados nas cidades do estado de São Paulo.
* Todos os anos que foram analisados os maiores índices de registros de feminicídio, ocorreram nas cidades de São Paulo e Campinas.

Figura 2: Quantidade de boletins por cidade.



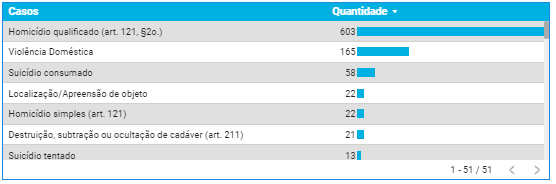
* **Porcentagem com os períodos de ocorrência**
* Gráfico apresenta a porcentagem dos períodos de ocorrência: manhã, tarde, noite, madrugada e em hora incerta.
* A análise mostra que o período da noite se encontra em primeiro lugar de registros (29,4 %), o período da manhã com 24,3 %, o período da tarde 23,1%, e de madrugada 19,9%, e em hora incerta apenas 3,3%.

Figura 3: Gráfico com o maior período de ocorrência.



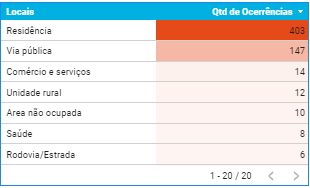
* **Casos de ocorrência:**
* Tabela com os casos registrados de boletins de ocorrência.
* Com a analise individual, os casos que mais ocorrem, são homicídio qualificado e violência doméstica.

Figura 4: Casos registrados por ocorrência.



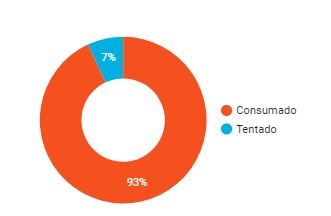
* **Quantidades de ocorrência por descrição do local do crime:**
* Tabela com o número de registros de boletins de ocorrência, com a descrição dos locais onde mais ocorrem o feminicídio.
* Com a analise individual de ano a ano, os locais ocorrem maiores índices do crime, são nas residências e via pública.

Figura 5: Quantidade de ocorrências por descrição do local.



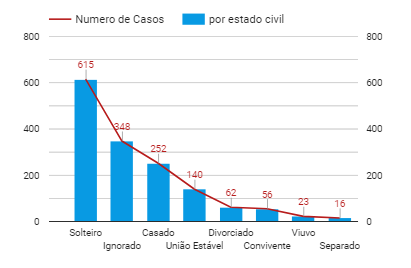
* **Porcentagem dos status dos boletins de ocorrência**
* Gráfico representando a porcentagem dos status registrados: consumado ou tentado.
* A análise mostra que a maioria dos casos foram “consumado” (93%), e apenas 7 % dos casos foram “tentado”.

Figura 6: Gráfico da porcentagem dos status dos boletins de ocorrência.



* **Número de casos relacionado ao estado civil da vítima:**
* Gráfico de barras combinado com linhas para demonstrar a contagem dos números de casos registrados, relacionado ao estado civil.
* A análise mostra que o maior número de casos identificados ocorre com pessoas com o estado civil de solteiro.

Figura 7: Gráfico com número de casos por estado civil.



* **Porcentagem de ocorrência por cor:**
* Gráfico com o registro em porcentagem de boletins de ocorrência, classificado quanto a cor: branca, parda, preta, não informado e outros.
* Com a analise individual a cor branca ocupa 47,6 % dos registros, parda 30,6%, não informada 15,6%, cor preta 5,7% e 0,5 % outros.

Figura 8: Gráfico com porcentagem de ocorrência por cor.

