**PROGRAMA DE MENTORIA ENTRE ALUNOS DSA**

**TURMA 2 - TEMPORADA 2021**

**Letícia Fiorina Echeverria Grobério-Mentorado**

**Luciana Silva Albuquerque de Melo-Mentorado**

**Rogélio Gonçalves Neto -Mentor**

**ANÁLISE DA CRIMINALIDADE NO ESTADO DE SÃO PAULO RELACIONADA A POPUPAÇÃO, EDUCAÇÃO E RENDA**

**BRASÍLIA**

**2021**

Sumário

[1 RESUMO 3](#_Toc83973512)

[2 OBJETIVOS 3](#_Toc83973513)

[3 MÉTODOS E FERRAMENTAS 3](#_Toc83973514)

[4 LIÇÕES APRENDIDAS 5](#_Toc83973515)

[4.1Os principais problemas encontrados no projeto 6](#_Toc83973516)

[4.2 Recomendações para melhorias futuras 6](#_Toc83973517)

[REFERÊNCIAS 7](#_Toc83973518)

# RESUMO

Para este Programa de Mentoria nossa equipe trabalhou em um projeto que teve como objetivo geral analisar de conjuntos de dados públicos sobre população, educação, renda e criminalidade no Estado de São Paulo a fim de encontrar possíveis correlações que possam corroborar com a tomada de decisão e planejamento estratégico por parte da Segurança Pública do Estado de São Paulo, visto a necessidade de maior agilidade no enfrentamento da criminalidade.

A fim de alcançar o objetivo principal do projeto a equipe seguiu alguns passos que estão listados como objetivos específicos e que propiciou atender, na medida do possível, o objetivo do Programa de Mentoria entre alunos DSA – Temporada - 2021, que é: “Permitir que alunos mais experientes e com perfil de Mentor, ajudem outros alunos indicando caminhos para os objetivos profissionais, guias de estudos, desenvolvimento de *soft skills* e trabalho em grupo” (DSA,2021).

Os conjuntos de dados foram extraídos por meio do desenvolvimento de um método automático e apresentados em um *dashboard* desenvolvido em Power BI.

# OBJETIVOS

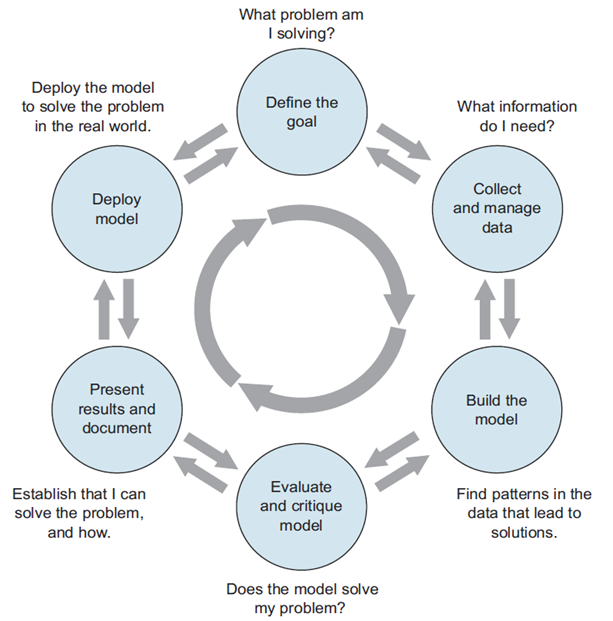
Este trabalho tem como objetivo geral a análise de conjuntos de dados públicos sobre população, educação, renda e criminalidade no Estado de São Paulo a fim de encontrar possíveis correlações que possam corroborar com a tomada de decisão e planejamento estratégico por parte da Segurança Pública do Estado de São Paulo e tem como objetivos específicos:

* Trabalhar com o ciclo de vida de ciência de dados;
* Extrair os quatro datasets propostos;
* Demonstrar os dados dos datasets extraídos;
* Apresentar um dashboard das informações;
* Apresentar um arquivo consolidado das informações;

# MÉTODOS E FERRAMENTAS

Parte deste trabalho transitou pelo ciclo de vida de um projeto de ciência de dados, conforme apresentado na figura 1. Cada fase será apresentada de maneira resumida.

Figura 1 – O ciclo de vida de um projeto de ciência de dados: loops dentro de loops



Fonte: (ZUMEL; MOUNT ,2014)

Na primeira fase é definido o objetivo, que no caso deste projeto é a análise de conjuntos de dados públicos sobre população, educação, renda e criminalidade no Estado de São Paulo a fim de encontrar possíveis correlações que possam corroborar com a tomada de decisão e planejamento estratégico por parte da Segurança Pública do Estado de São Paulo.

A segunda fase está relacionada aos dados, onde foram utilizados os datasets públicos das seguintes fontes:

|  |  |
| --- | --- |
| Datasets | Fonte |
| Escolas do Estado de São Paulo | http://www.educacao.sp.gov.br/central-de-atendimento/downloads.asp |
| Ocorrências Policiais do Estado de São Paulo | https://www.ssp.sp.gov.br/Estatistica/Pesquisa.aspx |
| PIB Municipal – Cidades do Estado de São Paulo | http://repositorio.seade.gov.br/dataset/pib-municipal-2002-2018 |
| População por Município, Sexo e Idade | https://repositorio.seade.gov.br/dataset/populacao-por-municipio-sexo-e-idade |

As linguagens de programação Python e R foram utilizadas na criação do método automático de extração dos datasets e para a criação do arquivo consolidado com os dados do projeto. Os códigos para extração dos dados para a montagem dos datasets foram criados no ambiente Jupyter Notebook com o uso da biblioteca Celenium. Para esta fase de pré-processamento dos dados estima-se que seja consumido pelo menos 70% do tempo total de um projeto de Data Science (ESCOVEDO; KOSHIYAMA,2020).

A terceira fase consiste na construção do método automático, ou seja, criação dos robôs extratores.

Na quarta fase foi realizada a avaliação, onde foi verificado se o método atendia de forma satisfatória ao que foi inicialmente proposto.

Em seguida, na quinta fase o projeto foi documentado, onde criou-se um repositório remoto no *Google Drive* para controle das versões de arquivos, armazenamento dos datasets e scripts. Este repositório viabilizou a gestão dos arquivos gerados ao longo de seu desenvolvimento, bem como a criação de um *dashboard* na ferramenta Power BI.

Seguindo o ciclo de desenvolvimento de um projeto de ciência de dados a última fase seria a de implementação, que não foi possível devido ao tempo que a equipe teve para a entrega dos trabalhos.

# LIÇÕES APRENDIDAS

Para tratarmos das lições aprendidas realizamos uma retrospectiva de como se deu o desenvolvimento do projeto deste Programa de Mentoria com o objetivo de corroborar na gestão do conhecimento bem como propiciar a melhoria contínua para projetos futuros em que os integrantes desta equipe venham participar.

No tocante as definições sobre lições aprendidas em projetos, destacamos algumas neste trabalho.

Para o Guia PMBOK (2013), lições aprendidas é o conhecimento adquirido durante um projeto, mostrando como os eventos foram abordados e como devem ser abordados no futuro para melhorar o seu desempenho.

Segundo a SBGC (Sociedade Brasileira de Gestão do Conhecimento) a lição aprendida é um conhecimento adquirido por meio da experiência em capturar, registrar, analisar e compartilhar.

Com o objetivo de evitar a sua recorrência, casos negativos e ou até mesmo o insucesso, impedir que um erro se repita em um determinado projeto ou atividade.

Promover sua aplicação em situações futuras e casos positivos, repetir um acerto nos próximos projetos ou atividades.

Em geral, as lições aprendidas podem ser consideradas como desvios em relação ao planejamento de um projeto ou atividade, tanto do ponto de vista positivo, como negativo.

A seguir serão apresentadas as lições aprendidas neste Programa de Mentoria.

# 4.1Os principais problemas encontrados no projeto

Um dos maiores problemas que encontramos no desenvolvimento deste projeto foi em conseguir reunir toda a equipe em todas as reuniões. Procurou-se agendar um dia e horário fixo na semana que todos tivessem disponibilidade, mas mesmo assim tivemos vários contratempos e por consequência o processo inicial de desenvolvimento se tornou moroso.

Destacamos também o pouco conhecimento técnico da equipe em relação as ferramentas que foram utilizadas. O integrante da equipe que tinha maior conhecimento em programação ficou sem disponibilidade para participar das reuniões e formalizou a desistência quase no final do projeto. O integrante com conhecimento em Power BI também teve alguns contratempos que impossibilitaram de participar das reuniões iniciais com maior efetividade e conseguiu contribuir com o projeto já na fase final.

# 4.2 Recomendações para melhorias futuras

Diante das dificuldades apresentadas, para projetos futuros recomendamos:

* Utilizar uma ferramenta online de gerenciamento de projetos;
* Incluir na pauta de todas as reuniões o item de lições aprendidas para que ao encerramento do projeto, fossem validadas pela equipe e desta forma proporcionar a melhoria contínua do desenvolvimento do projeto.

# REFERÊNCIAS

ESCOVEDO, T.; KOSHIYAMA, A. **Introdução a Data Science: algoritmos de machine learning e métodos de análise**. São Paulo: Casa do Código, 2020.

FUNDAÇÃO SEADE. **Pib Municipal 2002-2018.** São Paulo: SEAD,2021. Disponível em: <http://repositorio.seade.gov.br/dataset/pib-municipal-2002-2018>. Acesso em: 30 set. 2021.

FUNDAÇÃO SEADE. **População por município, sexo e idade.** São Paulo: SEAD,2021. Disponível em: < https://repositorio.seade.gov.br/dataset/populacao-por-municipio-sexo-e-idade>. Acesso em: 30 set. 2021.

PMBOK. **Um guia do conhecimento em gerenciamento de projetos** (Guia PMBOK) 5ª edição. Pennsylvania EUA: Project Management Institute, 2013.

DATA SCIENCE ACADEMY. **Programa de Mentoria entre alunos DSA temporada 2021**.Brasília: Blog DSA ,2021.Disponível em: <<https://blog.dsacademy.com.br/programa-de-mentoria-entre-os-alunos-dsa-temporada-2021/>>.Acesso em: 30 set. 2021.

SECRETARIA DE EDUCAÇÃO DO ESTADO DE SÃO PAULO. **Cadastro de escolas-Downloads.** São Paulo: SEDUC-SP,2021. Disponível em: <http://www.educacao.sp.gov.br/central-de-atendimento/downloads.asp>. Acesso em: 30 set. 2021.

SECRETARIA DE SEGURANÇA PÚBLICA DO ESTADO DE SÃO PAULO. **Dados Estatísticos do Estado de São Paulo.** São Paulo: SSP-SP,2021. Disponível em: < https://www.ssp.sp.gov.br/Estatistica/Pesquisa.aspx >. Acesso em: 30 set. 2021.

SBGC – Sociedade Brasileira de Gestão do Conhecimento. Disponível em: < http://www.sbgc.org.br/uploads/6/5/7/6/65766379/praticas\_de\_gc\_versao\_sbgc.pdf>, acesso em: 29 set. 2021.

ZUMEL N.; MOUNT, J. **Practical Data Science with R**. Shelter Island: Manning Publications,2014.