

Evolução do número de casos de contágios no Brasil, de acordo com diferentes variáveis demográficas

03 - Projeto

Germán Darío Buitrago Salazar *
Diego Braga Santi †
Christianne Orrico Dalforno *
Guilherme Aguilar de Oliveira †

1 Tema

O objetivo deste trabalho é realizar uma análise considerando diferentes variáveis demográficas como idade, etnia, sexo e escolaridade, correlacionando com a taxa de contágios de COVID-19 no Brasil. Esta análise irá detectar qual a frequência da ocorrência do COVID-19 em função de diferentes fatores demográficos e sua evolução desde o começo da pandemia. Ao finalizar a análise, pretende-se criar um modelo de predição de novos casos que relacione estes fatores.

2 Equipe

O nome da equipe é **Caçadores Demográficos**. O time está composto pelos membros a seguir.

Membro 1

Nome	Germán Darío Buitrago Salazar
RA	164321
Usuário no Gitlab	gedabusa
e-mail	gedabusa@gmail.com
Afiliação	Aluno especial de Doutorado em Ciência da Computação, UNICAMP

Membro 2

Nome	Diego Braga Santi
RA	2149613
Usuário no Gitlab	diegosanti
e-mail	diegobsanti@hotmail.com
Afiliação	Aluno de Doutorado - CPGEI, UTFPR

Membro 3

Nome	Christianne Orrico Dalforno
RA	233556
Usuário no Gitlab	codalforno
e-mail	profcod@gmail.com
Afiliação	Aluna especial, UNICAMP

*Universidade Estadual de Campinas, UNICAMP

†Universidade Tecnológica Federal do Paraná, UTFPR

Membro 4

Nome	Guilherme Aguilar de Oliveira
RA	2127954
Usuário no Gitlab	guilhermith
e-mail	guilhermeoliveira.2019@alunos.utfpr.edu.br
Afiliação	Sistemas de Informação, UTFPR

3 Perguntas de pesquisa

- Existe relação da escolaridade dos pacientes e a evolução de novos casos de COVID-19 ao longo do tempo?
- Há diferença na quantidade de casos em pessoas de diferentes etnias?
- Qual a relação entre o gênero e a incidência de infecção por COVID-19?
- Quantitativamente, como a COVID-19 tem afetado as diversas faixas etárias?
- Como o vírus da COVID-19 foi se espalhando geograficamente ao longo do tempo e qual seria sua relação com a densidade populacional da região?
- Como é possível relacionar os casos de COVID-19 com as variáveis descritas anteriormente?

4 Hipóteses

- Existe relação da escolaridade dos pacientes e a evolução de novos casos de COVID-19 ao longo do tempo?
 - Os casos iniciais ocorreram com maior frequência em pessoas com nível superior e em seguida o contágio atingiu as pessoas com outros níveis de escolaridade.
- Há uma incidência maior de casos de COVID-19 nas populações de etnia preta e parda em comparação às demais etnias.
 - Há uma incidência maior de casos de COVID-19 nas populações de etnia preta e parda em comparação às demais etnias.
- Qual a relação entre o gênero e a incidência de infecção por COVID-19?
 - Há maior incidência de casos de COVID-19 entre homens em comparação às mulheres.
- Quantitativamente, como a COVID-19 tem afetado as diversas faixas etárias?
 - Os casos de COVID-19 prevaleceram em pessoas na faixa etária de 30 a 50 anos com a maioria dos casos entre as pessoas na faixa de 30 a 40 anos.
- Como o vírus da COVID-19 foi se espalhando geograficamente ao longo do tempo e qual seria sua relação com a densidade populacional da região?
 - O crescimento de novos casos de COVID-19 foi mais acelerado em lugares com maior densidade populacional do que em lugares com menor número de pessoas.
 - Lugares com maior densidade populacional atingiram o pico de casos antes que lugares com menor densidade.
- Como é possível relacionar os casos de COVID-19 com as variáveis descritas anteriormente para definir o perfil dos infectados?
 - É possível definir classes de pessoas que possuem características similares e observar em quais delas a densidade de infectados é maior.

5 Dados e modelos

Com o objetivo de testar as hipóteses levantadas, algumas etapas no processamento de análise de dados serão realizadas. Na análise exploratória foram utilizadas duas bases de dados: SIVEP-gripe onde foram considerados dados de todos os estados do Brasil e e-SUS, inicialmente considerando somente os dados do Estado de São Paulo. Na próxima etapa todos os estados do Brasil serão analisados. Pretende-se utilizar somente os dados de pessoas que testaram positivo para COVID-19, ignorando os demais casos de SRAG. Dentro desse conjunto serão analisados e modelados os dados de sexo, etnia, escolaridade, idade e região. Uma análise individual de cada variável será realizada, verificando como foi a evolução de casos ao longo do tempo.

No caso das etnias, será levado em consideração o percentual de presença de cada grupo em cada região. Esses ajustes poderão dizer se determinadas etnias foram mais afetadas que outras, considerando o percentual de indivíduos que pertencem a determinadas etnias. O mesmo processo será feito para analisar a quantidade de casos por escolaridade, idade e sexo. Inicialmente uma análise geral, seguido de uma análise com ajustes necessários para cada região. Após análise individual, será realizado uma clusterização dessas variáveis para identificar grupos em que há maior ou menor predominância de infectados. Diferente das outras variáveis, para estudar a relação da densidade populacional com a quantidade de casos de COVID-19, pretende-se fazer uma regressão linear e identificar a possível correlação entre essas duas variáveis.

6 Atribuições dos membros júnior

Não haverá diferenciação de atribuições entre membros sênior e júnior, visto que o membro júnior tem conhecimento em ferramentas para análise dos dados.

7 Cronograma

Os próximos passos a serem realizados serão:

1. **Extração das bases de cada estado.** Para a análise exploratória, usou-se os dados do estado de São Paulo. Nesta etapa é realizada a extração da informação da base de dados disponibilizada.
2. **Limpeza e preparação de dados.** Os dados são preparados, removendo dados inconsistentes e selecionando a informação relevante para a pesquisa.
3. **União de datasets.** Os datasets de todos os estados são homogeneizados para conter a mesma estrutura de dados e juntados para formar uma única base de dados.
4. **Modelagem dos dados.** A base de dados se processa usando os modelos descritos.
5. **Ajuste de modelos.** A partir dos resultados obtidos na etapa anterior, são ajustados os modelos para melhorar sua performance.
6. **Revisão bibliográfica.** Estudo de outras pesquisas relacionadas com o foco de pesquisa que possam contribuir no trabalho.
7. **Análise de resultados finais e comparação com outros modelos.** Os resultados dos modelos são analisados e apresentados através de gráficos e realizada a comparação com resultados encontrados durante a revisão bibliográfica.
8. **Escrita do artigo final.** Documentação final do trabalho realizado em formato de artigo.

O cronograma está planejado por semana, considerando a primeira indo de 08/11/2020 a 14/11/2020 e a última de 27/12/2020 a 02/01/2021.

Tabela 1: Cronograma de atividades

Atividades	Responsável				Semanas							
	Germán	Diego	Christianne	Guilherme	01	02	03	04	05	06	07	08
1	X	X	X	X								
2	X	X	X	X								
3	X											
4		X		X								
5		X		X								
6	X		X									
7	X	X	X	X								
8	X	X	X	X								