

# Relatório Final de MACo213 - Atividade Curricular em Comunidade

**Aluno:** Luís Felipe de Melo Costa Silva  
**Número USP:** 9297961

## Introdução

Este projeto foi fruto de uma parceria entre o Tecs - Grupo de computação social da USP e a Secretaria Municipal de Educação (SME) de São Paulo. Ele consiste em uma Análise Histórica dos Dados sobre o Atendimento das Creches de São Paulo, de junho de 2006 a dezembro de 2017. Este relatório segue o que está escrito no [site](#) oficial da disciplina, que complementa esse texto com mais detalhes.

## 1 Tecnologias e Dados

Para a realização do projeto, foram usadas algumas fontes de dados. Uma delas foram os dados educacionais, provenientes do [site](#)<sup>1</sup> da prefeitura. Deles, apenas o recorte relacionado às creches (Matrículas e Demanda) foi utilizado.

Outra fonte foi a de dados populacionais, vindos do [site](#)<sup>2</sup> da Fundação SEADE. Desses dados, apenas a faixa etária de 0 a 3 anos foi utilizada. Junto com os dados educacionais selecionados, esses dados foram salvos num arquivo separado, utilizado para todas as análises.

Além desses dados, foram usados os dados das subdivisões de São Paulo em cinco zonas, de acordo com o Plano Diretor Estratégico de 2002 (PDE), também proveniente do [site](#) da Prefeitura.

O desenvolvimento da análise foi inteiramente feito em Python [3](#)<sup>3</sup>. A ferramenta utilizada para os códigos foi a Jupyter<sup>4</sup>, que coloca o desenvolvimento numa janela de navegador Web, como se vê na [Figura 1](#). A principal biblioteca de análise utilizada foi a Pandas<sup>5</sup>, que permite a conversão de dados .csv para estruturas de dados com usabilidade simples. Para os gráficos, foi usada a Matplotlib<sup>6</sup>.

---

<sup>1</sup><http://dados.prefeitura.sp.gov.br/>

<sup>2</sup><http://www.seade.gov.br/>

<sup>3</sup><https://www.python.org/>

<sup>4</sup><http://jupyter.org/>

<sup>5</sup><https://pandas.pydata.org/>

<sup>6</sup><https://matplotlib.org/>

## 1 - Filtro dos dados dos arquivos

Para que a análise seja mais eficiente, é necessário passar um certo filtro nos dados. O objetivo é, a partir dos dois arquivos de entrada (dados educacionais e dados populacionais), gerar apenas um arquivo que contenha apenas a informação necessária. Os dados das matrículas na rede municipal de educação vieram [daqui](#). Já os populacionais vieram do site da [SEADE](#), a partir de uma tabela nova.

```
In [4]: import pandas as pd

dados_educ = pd.read_csv('dados/evolucaodemanda.csv', sep = ';')
dados_pop = pd.read_csv('dados/populacao.csv', sep = ';', encoding = 'ISO-8859-1')

Para os dados educacionais, temos que filtrar apenas as colunas que necessárias. Além disso, é preciso limpar os dados, pois algumas linhas possuem os números com um ponto (.) para indicar o milhar. Por conveniência, o mês e o ano de referência serão deixados em colunas separadas.

In [5]: for i in dados_educ.index:
        if not dados_educ.at[i, 'MAT_CRECHE'].is_integer():
            dados_educ.at[i, 'MAT_CRECHE'] = round(dados_educ.at[i, 'MAT_CRECHE']*1000)
        if not dados_educ.at[i, 'DEM_CRECHE'].is_integer():
            dados_educ.at[i, 'DEM_CRECHE'] = round(dados_educ.at[i, 'DEM_CRECHE']*1000)
        dados_educ.at[i, 'MES'] = dados_educ.at[i, 'MES_ANO_REF'][:3]
        dados_educ.at[i, 'ANO'] = int(dados_educ.at[i, 'MES_ANO_REF'][4:]) + 2000

dados_educ = dados_educ[['DISTRITO', 'MAT_CRECHE', 'DEM_CRECHE', 'MES', 'ANO', 'MES_ANO_REF']]
```

**Figura 1:** Exemplo de arquivo no Jupyter.

## 2 Análises feitas

As análises geraram um **relatório** final, que compila todas as descobertas feitas durante o semestre. Ele está disponível *online*, e na íntegra.

### 2.1 Análise Municipal

A Análise Municipal juntou os dados dos distritos para obter os dados da cidade como um todo. Foi gerado um gráfico com a evolução da população total, das matrículas e da demanda, e, a partir disso, se tentou entender como tais variáveis se comportam. Se destaca a sazonalidade que acontece na fila e o crescimento quase que linear das matrículas. Também são vistas tabelas que comparam o primeiro e o último dados da amostra, assim como o primeiro e último mês de dezembro, para a demanda. Sua escrita levou **1 hora**.

### 2.2 Análise Distrital

A Análise Distrital segue o mesmo padrão da Análise Municipal. A diferença é que a unidade de análise são os distritos. A maioria dos distritos segue o mesmo padrão identificado no município. Foi feita aqui uma análise para ver quantos distritos estão numa situação confortável, ou seja, com mais matrículas do que demandas. Essa análise é acompanhada de um apêndice que possui uma página com mais informações para cada distrito. Sua escrita levou **8 horas**, pois, além de fazer a análise, foi necessário construir arquivos para gerar a página de cada distrito individualmente. Um exemplo de página gerada automaticamente está no [Apêndice A](#).

### 2.3 Distribuição do Atendimento

A terceira análise utiliza de mais recursos visuais para entender como o atendimento ocorre pelo município. São feitos mapas que analisam as matrículas e as demandas absolutas, as taxas de matrículas e demanda proporcionais à população. Também tem mapas que mostram a evolução do atendimento entre junho de 2006 e dezembro de

2017, para as matrículas e entre dezembro de 2006 e dezembro de 2017, para a demanda. Sua escrita levou **2h30**.

## 2.4 *Rankings*

Os *rankings* apresentam a classificação dos distritos em relação a quatro aspectos: matrículas absolutas, demanda absoluta, matrícula proporcional à população e demanda proporcional à população. Além disso, é feita uma comparação com um *ranking* do primeiro período comparativo da amostra (junho de 2006 para as matrículas e dezembro de 2006 para a demanda, por conta da sazonalidade da fila). Esse segundo *ranking* não está exibido no relatório final. O primeiro está anexo em um apêndice. Sua escrita levou **3h30**, já que, além de escrever sobre o que foi observado, foram gerados arquivos para elaborar as tabelas de *ranking*, que possuem 96 linhas cada, inviabilizando a escrita manual em tempo hábil.

## 2.5 Análise de Tendências

Aqui, tentou-se entender como se comportam três informações: o futuro das matrículas, o futuro das demandas e quando acontece uma inversão a favor das matrículas, ou seja, um distrito possui mais matrículas do que demanda. Para cada uma dessas informações existe um **ranking**, presente num apêndice do relatório. Sua escrita levou **3 horas**.

## 2.6 Relatório final

Além do tempo utilizado escrevendo sobre as análises, foi gasto certo tempo revisando o relatório e escolhendo o melhor estilo para ele, além de separar cada capítulo em um arquivo  $\text{\LaTeX}$ . Com isso, foram usadas **5 horas**. Portanto, as atividades relacionadas ao relatório tomaram **23 horas**.

# 3 Conclusão da Análise

A partir da análise dos dados, foi possível tirar algumas conclusões. Pôde-se ver que o município melhorou como um todo. Também se nota que ele está no caminho certo, com as matrículas aumentando e a fila diminuindo. No entanto, ainda faltam mais matrículas para que ele atinja a meta proporcional de 75%.

A evolução do atendimento foi diferente pelos distritos. Quase todos apresentaram aumento nas matrículas, uns mais do que outros, como o distrito de Lajeado, que subiu 14 posições no *ranking* de matrículas absolutas e 62 no *ranking* de matrículas proporcionais à população. Em contrapartida, o distrito da República foi o único que apresentou queda nas matrículas.

Pensando nas filas, é possível confirmar que elas caíram na maioria dos distritos. No entanto, alguns distritos apresentaram aumento, como a Vila Andrade, Pedreira e Marsilac.

O distrito do Marsilac, o mais meridional da cidade, é o único que atende mais de 100% de sua população estimada. Tal informação leva a uma reflexão quanto a acurácia das estimativas populacionais e também em relação a distritos que atendem crianças

de outro distrito, este último um fato que não pode ser ignorado ao longo da leitura dessa análise.

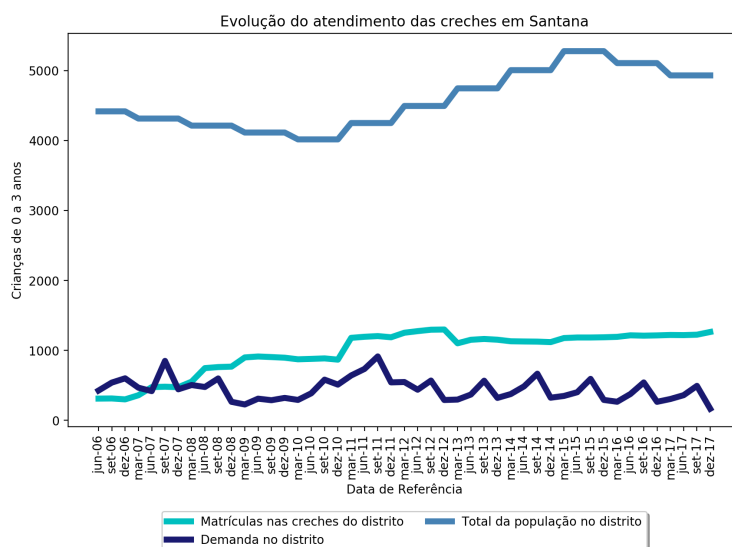
De acordo com os dados mais recentes, os distritos que apresentam o maior número de vagas se localizam na periferia das zonas Sul, Leste e Norte. São as mesmas regiões que concentram boa parte da demanda. Contudo, foram justamente essas regiões que tiveram os maiores aumentos no número de matrículas e maiores reduções na fila.

Portanto, conclui-se que, embora o município de São Paulo esteja melhorando no atendimento de creches, essa melhora não é refletida de maneira igual pela cidade. Isso não deve ser considerado um fato ruim, e se deve a diversas razões. Uma delas é que a população se distribui de maneira diferente pela cidade e outra é que nem todas as crianças se matriculam em creches municipais, frequentando a rede privada. Acontece também de crianças com idades entre 0 e 3 anos não estarem matriculadas em nenhum lugar, ficando em casa, com parentes, por exemplo.

## 4 Arquivos construídos

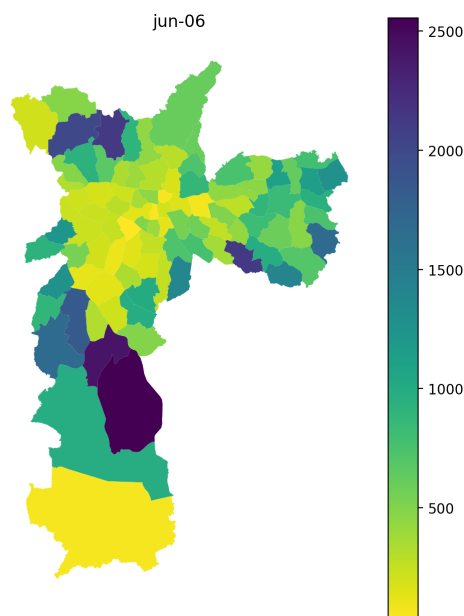
Para fazer as 5 análises supracitadas, foram gerados 10 arquivos em Python 3, cada um com uma parte da análise.

1. **Filtro de arquivos:** Sua ideia era de pegar os arquivos de dados obtidos das fontes externas, limpar o que não seria usado durante o semestre, arrumar parte dos dados e então, salvar o produto em um arquivo resultante que seria usado em todos os outros *notebooks*. Contribui com **todas** as análises. Seu desenvolvimento levou **7 horas**.
2. **Gráficos de tempo:** Gera um gráfico de evolução da população, das matrículas e da demanda para cada um dos 96 distritos da cidade de São Paulo, a partir dos dados filtrados. Um exemplo de seu produto pode ser visto na [Figura 2](#). Esse arquivo ajudou na elaboração da **Análise Distrital** e da **Análise Municipal**, pelo conhecimento obtido. Seu desenvolvimento levou **5 horas**.



**Figura 2:** Evolução das matrículas, da demanda, e da população de 0 a 3 anos no distrito Santana.

3. **Mapas:** Esse *notebook* serve para gerar os mapas de distribuição das matrículas e da demanda pela cidade. Foi o que levou mais tempo para ficar pronto, em torno de 10 dias. Um de seus resultados pode ser visto na [Figura 3](#). Ele contribuiu com o desenvolvimento da análise de **Distribuição do Atendimento**. Seu desenvolvimento levou **17 horas**.



**Figura 3:** Mapa que mostra o número de matrículas pela cidade em junho de 2006, primeiro dado da amostra.

4. **Análise Municipal:** Cuida de transformar (no caso, somar) todos os dados referentes ao município (total populacional estimado de 0 a 3 anos, número de matrículas e demanda) para posterior análise. Ajuda na **Análise Municipal**, como pode-se pensar. Também gera um gráfico como o da [Figura 3](#). Seu desenvolvimento levou **2 horas**.
5. **Agrupamentos:** A ideia de criar esse arquivo era para gerar gráficos de dispersão que mostrariam como os distritos podem ser agrupados, no entanto, após algumas experiências, foi observado que os distritos sempre estão relativamente próximos, em taxas de atendimento relativas. Gráficos feitos aqui podem ser visto na análise de **Distribuição do Atendimento**. Seu desenvolvimento levou **3h30**.
6. **Comparação:** Esse *notebook* serviu para comparar cada distrito com o município. Foram feitos gráficos que mostram a evolução da participação proporcional de cada distrito no município (um deles pode ser visto no *pôster* feito para a disciplina) e gráficos que exibem a evolução do distrito no *ranking* municipal (esses últimos nunca usados, porém essenciais para a ideia da análise de **Rankings**). Seu desenvolvimento levou **3h30**.
7. **Listas:** Constrói dois gráficos apenas: um que mostra o número de matrículas em junho de 2006 nos distritos sobreposto às matrículas em dezembro de 2007 e um outro que compara a demanda em dezembro de 2006 com a demanda em dezembro de 2007. O primeiro foi utilizado no *pôster* feito para a disciplina e o

outro está presente numa versão *alternativa* do pôster. O conhecimento obtido aqui foi utilizado para a análise de **Rankings**. Seu desenvolvimento levou **8h30**.

8. **Previsão da demanda:** Usando um modelo conhecido como ARIMA<sup>7</sup>, foi construída uma previsão de como a demanda vai se comportar num futuro próximo, a partir dos dados de entrada. Com isso, foi feita uma estimativa que mostra quando a demanda atingirá o valor zero. Usada na **Análise de Tendências**. Seu desenvolvimento levou **10 horas**.
9. **Previsão das matrículas:** Funciona do mesmo modo que o *notebook* anterior. Sua variação se deve pelo fato de ele ser feito para a previsão das matrículas. A meta aqui não é atingir um valor absoluto, porque os distritos não têm as mesmas populações. Por isso, foi usado um valor proporcional de 75%. Como não existem dados de estimativas populacionais para anos depois de 2018, foi usada a média populacional entre 1995 e 2017 como parâmetro de comparação. Mais detalhes estão na **Análise de Tendências**. Seu desenvolvimento levou **5 horas**.
10. **Inversões:** Procura quando ocorreu a última inversão entre demandas e matrículas, que é quando o número de matrículas fica maior que a demanda. Assume-se que uma posição confortável para um distrito seja quando ele possui mais matrículas do que demandas. Um total de 63 distritos possuíam mais fila do que matrículas em junho de 2006. Já em dezembro de 2017, apenas a Sé encontrava-se nessa condição. Para a **Análise de Tendências**, foi gerada uma tabela que mostra quando as inversões aconteceram pelos distritos. Seu desenvolvimento levou **2 horas**.

Também fiz uso de um tempo para padronizar a geração de gráficos no geral, buscar outros dados, como renda *per capita*, e modularizar e revisar os arquivos. Isso tudo levou **5 horas**. Portanto, com o desenvolvimento dos arquivos da análise foram usadas **68h30**.

## 5 Relação com a SME

A pessoa da SME com quem mais tive contato era a minha supervisora, Priscilla Corrêa, Assessora Técnica na Secretaria Municipal de Educação de São Paulo. Também tive contato com a Ananda Grinkraut, Analista de Políticas Públicas e Gestão Governamental. Nossa relação era pautada por conversas por e-mail e por mensagens. Uma foto nossa pode ser vista na [Figura 4](#)

Fizemos três reuniões presenciais, duas delas na sede do Pátio Digital, na casa Sérgio Buarque de Holanda, recentemente reformada, localizada no Pacaembu, na zona Oeste da cidade. Uma outra reunião foi feita na sede da SME, na Vila Clementino, zona Sul da cidade.

A primeira reunião, de apresentação da SME e do projeto, foi em 30 de agosto e durou **1 hora**.

A segunda reunião, de acompanhamento e apresentação aconteceu em 27 de agosto, e também durou **1 hora**. A preparação para ela levou **6 horas**.

Haveria uma reunião em outubro. Para ela, eu havia planejado apresentar um *design* de páginas para fazer o relatório *online*. Sua execução levou **8h30**, incorretamente

---

<sup>7</sup><https://pt.wikipedia.org/wiki/ARIMA>

contados como tempo gasto com análises, na versão prévia deste documento. Essa reunião não aconteceu, e as páginas foram substituídas pelo relatório em formato pdf.

A última reunião ocorreu no dia 30 de novembro, e levou **1 hora**. Nela, foram apresentados os últimos resultados das análises, e foi a única na sede da SME.



**Figura 4:** Da esquerda para a direita, Priscilla, eu e Ananda.

Portanto, com as reuniões e o preparo para elas foram utilizadas **17h30**.

## 6 Horas trabalhadas

Para esta disciplina foi usado um total de **109 horas**, passando o total de 100 horas estipulado.

O tempo usado para a matéria foi dividido basicamente em três partes: Reuniões e Apresentações com a SME, Análises de Dados e Elaboração do Relatório Final. Os detalhes sobre as horas gastas estão espalhados por este documento. Em resumo:

<b>Atividade</b>	<b>Horas Registradas</b>
<i>Reuniões e Apresentações com a SME</i>	17,5
<i>Análises</i>	68,5
<i>Elaboração do Relatório Final</i>	23

A relação hora-mês pode ser vista na tabela a seguir:

<b>Atividade</b>	<b>Horas Registradas</b>
<i>Agosto e Setembro</i>	25
<i>Outubro</i>	40
<i>Novembro</i>	44



Não estão incluídos nesses registros os tempos de elaboração do pôster da disciplina e da produção desse relatório. Também foram excluídos os tempos de comunicação com a SME (por e-mail e por mensagem) e tempos de deslocamento para as reuniões.

Mais detalhes podem ser vistos na página de **estatísticas** do site da disciplina.

No **Apêndice B**, está a confirmação das 100 horas trabalhadas.

## **7 Parte Subjetiva e Projetos Futuros**

Na minha opinião, o projeto foi muito enriquecedor. Foi a primeira vez que fiz um projeto grande de análise de dados. Não tive muita facilidade no começo, mas depois tudo começou a fazer mais sentido.

Foi a primeira vez que tive algum contato com um órgão governamental. A concepção muda bastante ao conhecer o trabalho que é feito por eles ao invés de só ouvir as notícias, que costumam destacar fatos negativos.

Para essa matéria precisei de muita organização. O primeiro motivo para isso é que os resultados seriam entregues apenas no fim do semestre. Quando vi que a média de horas que eu dedicava à disciplina estava aquém do necessário para cumprir as 100 horas mínimas, tive que estabelecer uma rotina de trabalhar para as análises pelo em torno de 90 minutos por dia. É possível observar que isso funcionou.

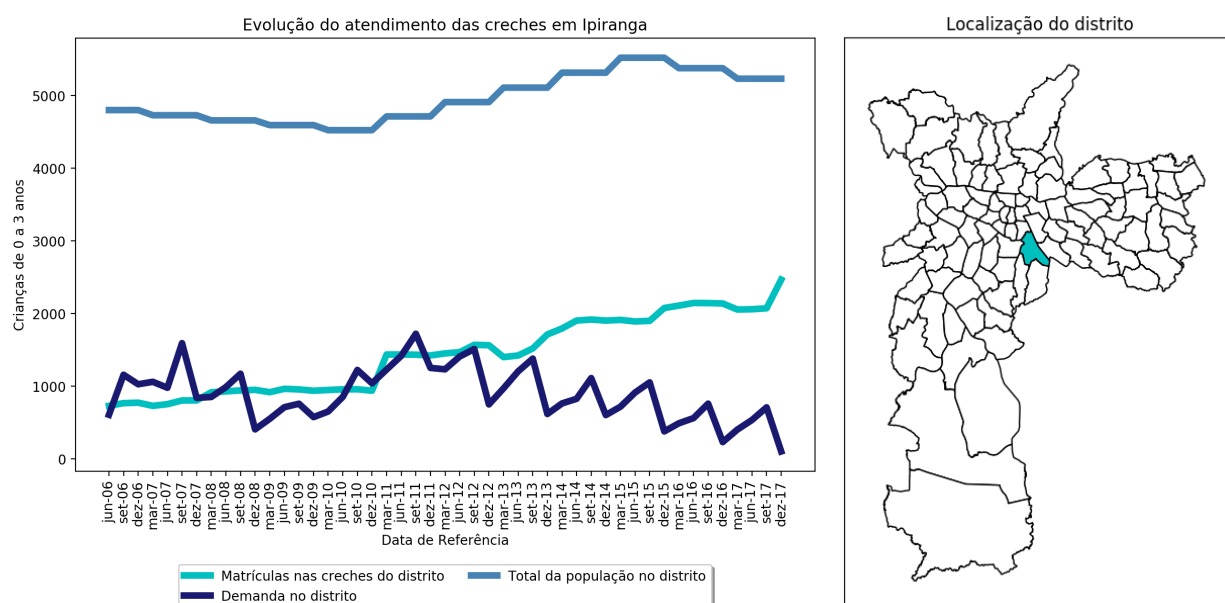
As análises feitas aqui foram bem simples, e, no final, não precisaram de tanto conhecimento técnico. Para as próximas análises nesse tema, pode ser estudado quem são as crianças que frequentam as creches municipais, traçando um perfil sócioeconômico dos estudantes. Também é possível analisar quais são os critérios que mais influenciam na escolha das creches municipais. Algumas variáveis podem ser pesquisadas, tais como oferta de vagas no distrito e renda das famílias. É possível também procurar agrupamentos entre os distritos, para caracterizá-los e traçar ações em comum para os grupos eventualmente identificados.



## **A Análise Distrital**

Exemplo de página gerada automaticamente sobre a análise de um distrito. Está na página a seguir.

### A.33 Ipiranga



**Figura A.33:** Evolução das matrículas, da demanda, e da população de 0 a 3 anos no distrito Ipiranga e a sua localização no município, em azul-claro, respectivamente.

	junho de 2006	dezembro de 2017	Variação
<b>Número de matrículas</b>	728	2465	238.6%
<b>Crianças na fila</b>	605	100	-83.47%
<b>População de 0 a 3 anos estimada</b>	4800	5232	9.0%
<b>Matrículas relativas à população</b>	15.17%	47.11%	210.64%
<b>Fila relativa à população</b>	12.6%	1.91%	-84.84%
<b>Proporção das matrículas no município</b>	1.18%	0.83%	-29.45%
<b>Proporção da demanda no município</b>	0.72%	0.23%	-68.36%
<b>Proporção da população de 0 a 3 anos</b>	0.76%	0.81%	6.2%
<b>Posição no ranking das matrículas</b>	33	45	-12
<b>Posição no ranking da demanda</b>	45	83	-38

**Tabela A.65:** Comparação entre o primeiro e o último período da amostra

	dezembro de 2006	dezembro de 2017	Variação
<b>Crianças na fila</b>	1025	100	-90.24%
<b>População de 0 a 3 anos estimada</b>	4800	5232	9.0%
<b>Fila relativa à população</b>	16.06%	1.91%	-88.1%

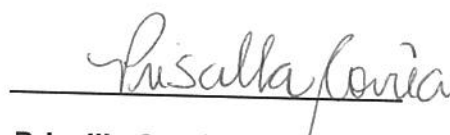
**Tabela A.66:** Comparação da demanda no mês de dezembro, em 2006 e 2017.

## **B Declaração de Horas Trabalhadas**

Declaração, feita para a SME, confirmando o total mínimo de horas trabalhadas para a disciplina. Segue na próxima página.

## DECLARAÇÃO

Declaro para os devidos fins que o aluno **LUÍS FELIPE DE MELO** (Nº USP: 9297961), regularmente inscrito na disciplina MAC0213 - "*Atividade Curricular em Comunidade*" do Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP), cumpriu junto a esta Instituição 100 horas de atividades, durante o 2º semestre de 2018.



**Priscilla Corrêa dos Santos**  
Secretaria Municipal de Educação  
*Assessora Técnica*  
R.F.: 839.189.1