

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 em no [site](#) oficial da disciplina, que complementa esse texto com mais detalhes.

1 Tecnologias e Dados

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

Outra fonte foi a de dados populacionais, vindos do [site](#)² 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.

O desenvolvimento da análise foi inteiramente feito em Python 3³. A ferramenta utilizada para os códigos foi a Jupyter⁴, que coloca o desenvolvimento numa janela de navegador Web. A principal biblioteca de análise utilizada foi a Pandas⁵, que permite a conversão de dados .csv para estruturas de dados com usabilidade simples. Para os gráficos, foi usada a Matplotlib⁶.

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.

¹<http://dados.prefeitura.sp.gov.br/>

²<http://www.seade.gov.br/>

³<https://www.python.org/>

⁴<http://jupyter.org/>

⁵<https://pandas.pydata.org/>

⁶<https://matplotlib.org/>

2.1 Análise Municipal

2.2 Análise Distrital

2.3 Distribuição do Atendimento

2.4 Rankings

2.5 Análise de Tendências

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 do consumo desta 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. No entanto, 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. Uma outra é que nem todas as crianças se matriculam em creches municipais, e sim, na 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 uma 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.

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 1](#). Esse arquivo ajudou na elaboração da **Análise Distrital** e da **Análise Municipal**, pelo conhecimento obtido.

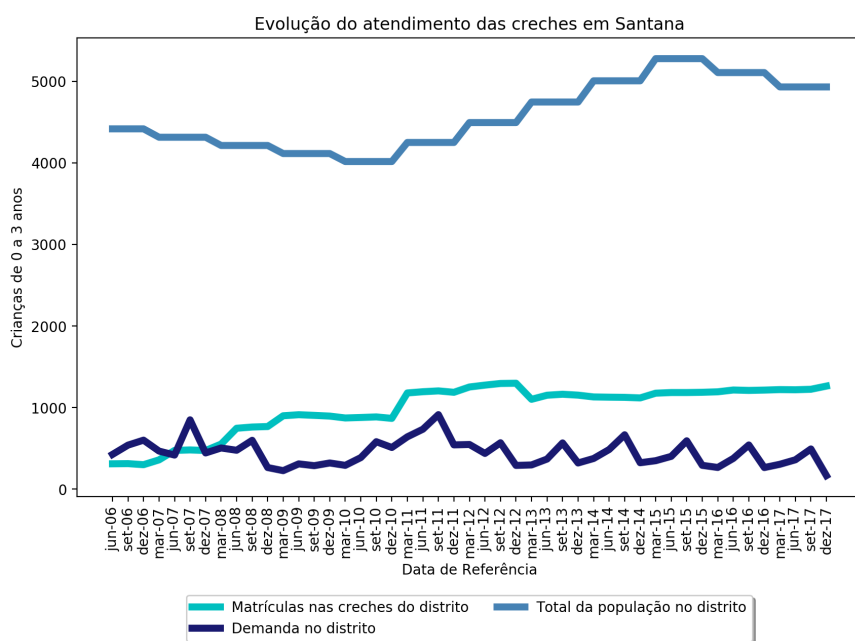


Figura 1: Evolução das matrículas, da demanda, e da população de 0 a 3 anos no distrito Santana.

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 2](#). Ele contribuiu com o desenvolvimento da análise de **Distribuição do Atendimento**.
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 2](#).
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**.
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**).

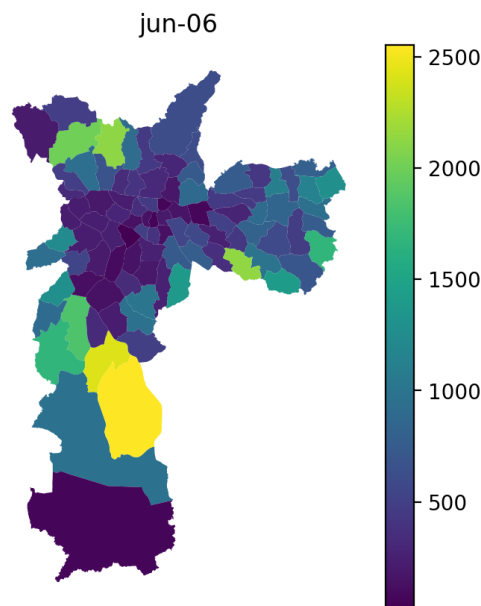


Figura 2: Mapa que mostra o número de matrículas pela cidade em junho de 2006, primeiro dado da amostra.

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**.
8. **Previsão da demanda:** Usando um modelo conhecido como ARIMA⁷, 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**.
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 ficam na **Análise de Tendências**.
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 matrículas em junho de 2006. Já em dezembro de 2017, apenas a Sé se encontra nessa condição. Para a **Análise de Tendências**, foi gerada uma tabela que mostra quando as inversões aconteceram pelos distritos.

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

Também foram feitos alguns *notebooks* auxiliares para a produção de texto automatizado para o Relatório Final entregue à SME.

5 Relação com a SME

6 Horas trabalhadas

Para esta disciplina, foi usado um total de **108 horas**, divididas como na tabela abaixo:

Atividade	Horas Registradas
<i>Reuniões e Apresentações com a SME</i>	8
<i>Análises</i>	77
<i>Elaboração do Relatório Final</i>	23

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

Atividade	Horas Registradas
<i>Agosto e Setembro</i>	25
<i>Outubro</i>	40
<i>Novembro</i>	43

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.

7 Conclusão do Projeto e Parte Subjetiva