

Relatório Final - MAC213

Projeto : Desenvolvimento de tutoriais de manuseio e análise de dados

Guilherme Yambanis Thomaz - 8041265

Objetivos

Com o intuito de fortalecer a transparência, a SME tem publicado e atualizado um conjunto de bases de dados referentes à educação municipal, variando de planilhas a microdados. O acesso a tais bases de dados pode ser, muitas vezes, limitado, devido à falta de conhecimento para manuseio das bases do público interessado (jornalistas, pesquisadores, integrantes de conselhos escolares). Frente a esse desafio, o Pátio Digital lançou o “ABC Dados”, que consiste na publicação de tutoriais que visam a orientar, numa linguagem didática e fluida, metodologias de manuseio das bases.

Tarefas

Após reunião realizada no dia 31 de agosto com Priscilla, na sede do Pátio Digital, definiu-se que os temas dos tutoriais seriam Excel e Python, e que o controle de horas seria feito através da ferramenta Toggl.

Os dez tutoriais a serem realizados ficaram igualmente divididos entre Excel e Python.

Os cinco tutoriais de cada plataforma foram estruturados da seguinte forma:

Tutoriais Excel

Tutoriais acessíveis, sobre o Excel, para manipulação de dados, em especial arquivos csv.

Tutorial 1

- O que é uma planilha?
- Morfologia básica
- Importando conjunto de dados

Tutorial 2

- Introdução a funções
- Função SOMA
- Função CONT.SE
- Função SE

- Função SOMASE

Tutorial 3

- Criando um gráfico de pizza simples
- Customizando o gráfico

Tutorial 4

- Google Sheets
- Importando csv
- Usando funções no Google Sheets
- Criando uma tabela dinâmica no Google Sheets
- Criando um gráfico a partir dos dados
- Função FREQUÊNCIA

Tutorial 5

- Criando uma tabela Dinâmica
- Manipulação e visualização dos dados

Tutoriais Python

Tutoriais criado na plataforma Jupyter Notebook.

Tópicos abordados:

Tutorial 1

- Noções elementares de Python

Tutorial 2

- Pacote Pandas

Tutorial 3 e 4

- Criando gráficos com Matplotlib

Tutorial 4

- Regex

Tutorial 5

- Definindo funções

Progresso e desafios

Excel

Defini inicialmente que os tutoriais seriam escritos em Markdown, por ser um formato de fácil exportação, seja para plataforma web, seja para pdf, além da possibilidade de inserção de links e gifs.

A construção e definição dos tutoriais tomou um certo tempo, porém, o maior desafio se mostrou na criação dos gifs, que demandaram diversas tentativas com ferramentas diferentes.

Além disso, após a parte técnica finalizada, escrever o texto de forma enxuta e didática também se mostrou algo desafiador. Grande parte da carga horária foi dedicada para refacção e reescrita, a fim de maior clareza e melhor didática.

Como o Excel depende da interação com a interface gráfica, foi necessário que o texto fosse bastante descritivo. Os gifs ajudaram muito nesse sentido, comunicando ações mais fáceis de serem traduzidas de maneira visual.

Python

Já para os tutoriais em Python, preferi o formato Notebook.

Iniciei os tutoriais usando a plataforma do Google, o Colab. Entretanto, após algum tempo, encontrei diversas dificuldades, como importações incompletas entre outras.

Isso tornou necessário migrar o projeto para o formato Jupyter Notebook. Como ele também é renderizado no próprio Github, não se mostrou uma perda muito grande nesse aspecto, além do ganho em robustez pela execução local do código.

A criação dos tutoriais em Python se mostrou mais desafiadora do ponto de vista técnico, exigindo muitas refacções e pesquisa sobre a linguagem.

Certas tarefas que propus nos tutoriais se mostraram mais complexas do que o previsto, mas isso permitiu não só que eu aprimorasse meus conhecimentos de Python, como também que eu criasse os tutoriais de um ponto de vista mais empático com o leitor, o que, acredito, melhorou a didática.

A junção do código com o texto em um só lugar, graças ao Jupyter Notebook, permitiu a criação de um tutorial muito mais intuitivo, já que o código e sua explicação estão lado a lado e o aluno pode executá-lo em tempo real.

Resultado

O resultado final foram 10 tutoriais, que servem para apresentar as duas plataformas, Python e Excel, de maneira amigável.

O intuito foi partir de um ponto inicial bastante básico, para que, independente do nível de conhecimento do leitor, fosse possível a utilização das ferramentas, sem uma barreira de entrada alta e intimidadora.

Os tutoriais iniciais são muito explicativos, sem pressupor nenhum conhecimento prévio.

Após os tutoriais iniciais de cada assunto, temas mais complexos começaram a ser abordados de forma mais pratica, o que foi possivel gracias a introdução aos *features* anteriormente.

É esperado que, uma pessoa que acompanhe todos os tutoriais, seja capaz de sair de um estado de total desconhecimento sobre as ferramentas e possa criar análises de dados pertinentes com as técnicas adquiridas ou utilize-os de ponto de partida para procurar informações mais específicas, seja pro Python ou Excel, já tem certa familiaridades com as plataformas.

Poster

O poster foi fixado no IME no dia 8 de novembro.

aluno responsável
GUILHERME YAMBANIS THOMAZ
supervisora
PRISCILLA CORRÊA, ASSESSORA TÉCNICA DA SME
disciplina
MAC 213 - Atividade Curricular em Comunidade
repositório
<https://github.com/yambanis/MAC213>

DESENVOLVIMENTO DE TUTORIAIS DE MANUSEIO E ANÁLISE DE DADOS

O QUE É O PÁTIO DIGITAL?

POLÍTICA DE GOVERNO ABERTO DA PREFEITURA
DE SÃO PAULO, ESTRUTURADA EM TRÊS EIXOS:

1. TRANSPARÊNCIA E DADOS ABERTOS
2. INOVAÇÃO TECNOLÓGICA
3. COLABORAÇÃO GOVERNO-SOCIEDADE

PROJETO

O PÁTIO DIGITAL LANÇOU O "ABC DADOS": TUTORIAIS QUE VISAM A ORIENTAR, NUMA LINGUAGEM DIDÁTICA E FLUIDA, METODOLOGIAS DE MANUSEIO DAS BASES DE DADOS REFERENTES À EDUCAÇÃO MUNICIPAL

DURANTE O SEMESTRE, FORAM DESENVOLVIDOS TUTORIAIS EM EXCEL E PYTHON, VISANDO A AUXILIAR, NO ACESSO E NA MANIPULAÇÃO DESSAS BASES DE DADOS, O PÚBLICO INTERESSADO QUE NÃO CONTA COM O CONHECIMENTO NECESSÁRIO PARA O SEU MANUSEIO

EXCEL

TUTORIAIS DESENVOLVIDOS EM
MARKDOWN, COM UTILIZAÇÃO
DE GIFS, PARA MELHOR DIDÁTICA

PYTHON

TUTORIAIS DESENVOLVIDOS EM
NOTEBOOK, PARA INCORPORAR
CÓDIGO E TEXTO EM UM ÚNICO
ARQUIVO

Acompanhamento

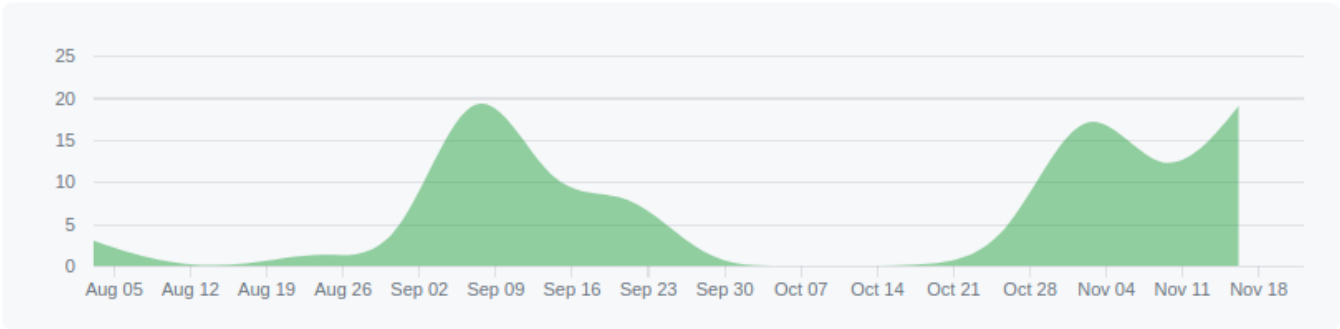
O acompanhamento se deu de duas formas principais:

- [Repositorio](#) aberto no Github, onde todas os arquivos foram salvos durante o semestre

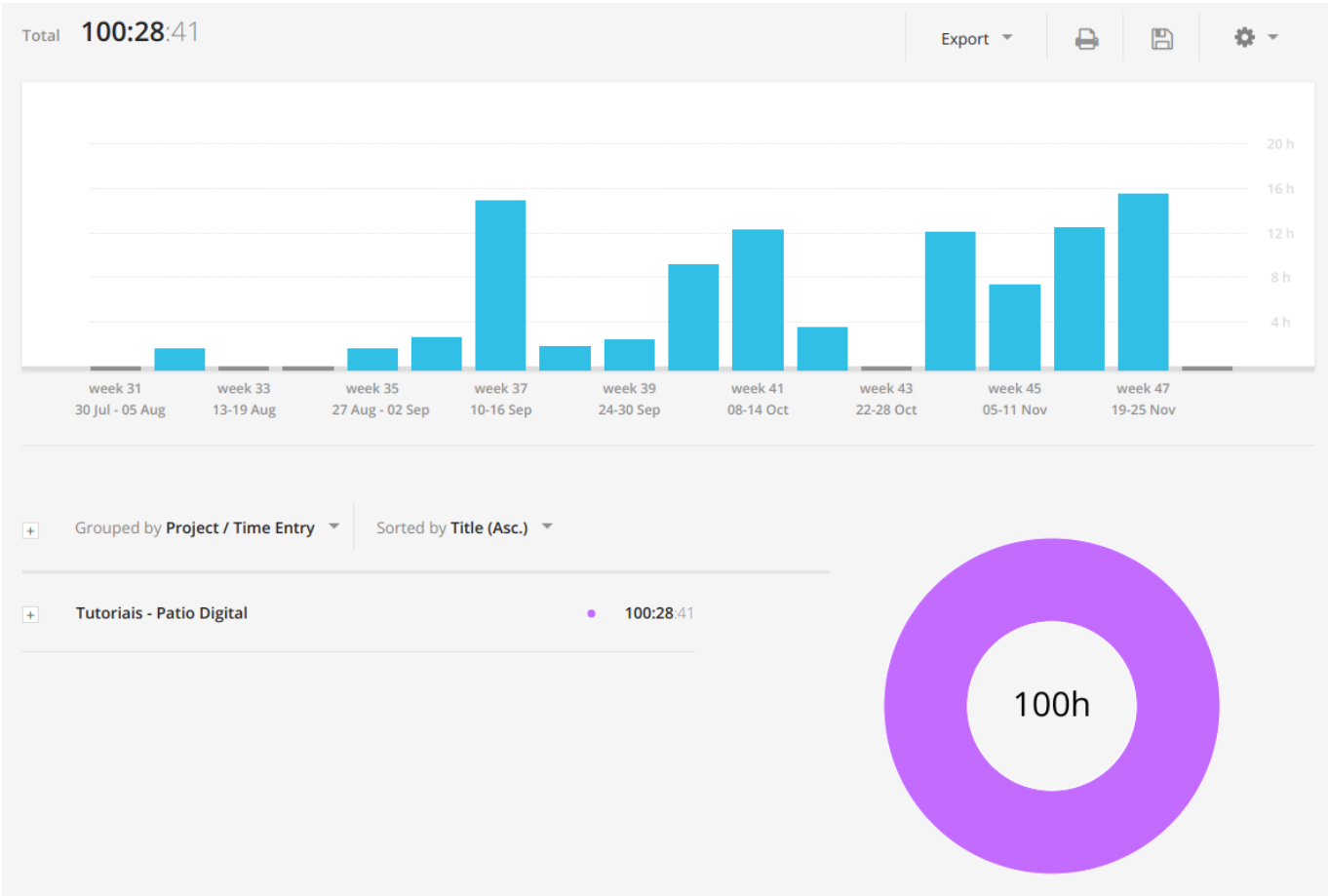
Aug 5, 2018 – Nov 24, 2018

Contributions: **Commits** ▾

Contributions to master, excluding merge commits

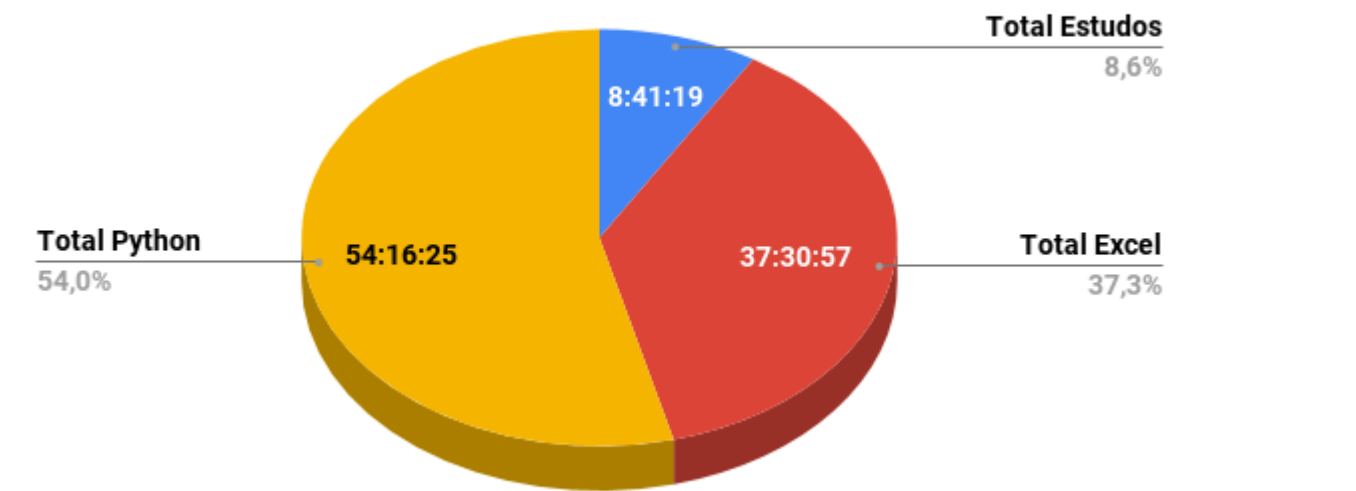


- Controle de horas realizado através da ferramenta Toggl



Horas despendidas no projeto

Duração das Atividades(24/11)



Horas despendidas (detalhado)

Estudos	Duracao	Total Estudos	Excel	Duracao	Total Excel	Python	Duracao	Total Python
Criacao do Github, estudos iniciais	1:56:00	8:41:19	Inicio tutorial 1	1:55:35	37:30:57	planejamento Python	0:55:28	54:16:25
Reuniao Patio Digital	2:00:00		Tutorial Excel	1:59:37		inicio tutorial Python	2:24:33	
Pesquisas Iniciais - Topicos Excel	1:32:00		Tutorial funcoes	0:15:58		Continuacao tutorial Pandas	2:08:04	
Pesquisando Referencias - Tutoriais	1:29:00		Tutorial de formulas	2:16:33		Mais Pandas	2:12:00	
Pesquisando Topicos	1:44:19		tutorial GSheets	0:58:46		migrando para Jupyter	3:40:24	
			revisões	1:49:26		Python Md	4:12:17	
			Revisões	0:53:27		python	2:20:07	
TOTAL GERAL	100:28:41		correcao orto	1:31:50		Tutorial Python - continuação	2:17:48	
			checagem	0:29:59		Palestra Patio Digital	1:29:44	
			Refatorando	0:58:34		Cartaz	2:03:18	
			Refatorando	1:13:58		matplotlib	2:58:00	
			revisoes	0:57:23		Revisao Cartaz	1:05:07	
			arrumando gifs e tutorial funcoes	1:56:08		Python Grafico de barras	0:38:09	
			Poster	6:00:28		matplotlib	0:57:58	
			correcoes ortográficas e revisoes	4:53:55		Revisao Final Poster	1:30:04	
			Revisao CountSe Excel	2:15:55		impressão poster	0:58:00	
			grafico de barras excel	1:17:40		revisao Python	0:29:19	
			Tutorial Tabela Dinamica	0:53:04		matplot Bar	0:00:10	
			Continuacao tutorial Tabela Dinamica	2:10:41		matplot Bar	3:13:50	
			Correcoes Finais	2:42:00		matplot Bar	2:16:21	
						Regex and Groupby	2:05:43	
						Modularizando Tutoriais	2:33:21	
						Correcoes Ortograficas	2:00:48	
						Mais correcoes	3:29:17	
						Revisao de Texto	2:49:31	
						Revisoes fianis Python	1:20:00	
						Exportar Python	0:21:27	
						Novo Tutorial - funcoes	1:45:37	