**PROJETO SEDURB**

Autor:

Matheus E. S. Azevedo (Analista de Dados)

|  |  |  |
| --- | --- | --- |
| **NOME** | **SEÇÃO** | **DATA** |
| Matheus E S Azevedo |  | 20180606 |
| Matheus E S Azevedo |  | 20180608 |
| Matheus E S Azevedo |  | 20180609 |
| Matheus E S Azevedo |  | 20180611 |
|  |  |  |

**REGISTRO DE ALTERAÇÕES**

**Sumário**

1. **O PROJETO SEDURB1**
2. **DADOS ORIGINAIS4**

**201806064**

1. **PREPARAÇÃO DOS DADOS4**

**201806064**

1. **TRATAMENTO DOS DADOS4**

**201806064**

**201806084**

**201806114**

1. **INSIGHTS4**

**201806064**

**201806084**

**201806094**

1. **ARQUIVO FINAL4**

**201806094**

1. **CONCLUSÃO4**
2. **O PROEJTO SEDURB**

O Projeto SEDURB (Secretaria de desenvolvimento Urbano) trata-se de uma lista de candidatos classificados para a segunda fase do concurso organizado pelo órgão municipal de João Pessoa.

A lista de candidatos contém três colunas (Inscrição, Candidato e Nota Objetiva) e esta ordenada por ordem alfabética, este ultimo fato foi o que gerou a necessidade de trabalharmos com esta lista.

O objetivo deste projeto é ordenar a lista de candidatos de acordo com as notas, permitindo que estes possam consultar a lista e verificar se foram aprovados para próxima fase do concurso. Outro detalhe importante contido no edital está relacionado as vagas disponibilizadas pelo órgão municipal, pois apenas 266 candidatos estarão aptos para a fase seguinte do concurso.

* 1. Detalhes Técnicos

O projeto será desenvolvido utilizando o ambiente de desenvolvimento Jupyter Notebook, por se tratar de uma ferramenta bastante eficiente para demonstrar os resultados ao longo do projeto.

Já a estrutura de organização do projeto será dívida em etapas que refletem as boas práticas do tratamento de dados que são:

> Dados Originais;

> Preparação dos Dados;

> Tratamento dos Dados;

> Insights;

> Arquivo Final.

Em seguida cada etapa conterá os arquivos que serão manipulados e subdivididos por data, dessa forma sendo possível visualizar cada modificação realizada no projeto.

Por fim, cada alteração será registrada no Registro de Alterações possibilitando a identificação imediata da modificação e do autor da mesma.

1. **DADOS ORIGINAIS**

Nesta etapa recebemos os dados originais do projeto que por segurança permaneceram inalterados até o fim do mesmo, os dados serão catalogados conforme a data em que foram recebidos.

20180606

Foi recebido um PDF (Resultado-preliminar-dos-Candidatos-Aprovados-OBJ) do certame da SEDURB, contendo os dados que serão manipulados no projeto referente aos candidatos à vaga de Agente de Controle Urbano.

Este arquivo é uma lista contendo os campos:

> Inscrição (Número de sua inscrição no certame);

> Candidato (Nome do candidato);

> Nota Objetiva.

1. **PREPARAÇÃO DOS DADOS**

Nesta etapa realizamos a cópia dos dados originais contidos na etapa anterior do projeto aonde receberão o primeiro tratamento, os dados serão catalogados conforme a data em que foram recebidos.

20180606

Foi recebido um PDF (Resultado-preliminar-dos-Candidatos-Aprovados-OBJ) do certame da SEDURB, contendo os dados que serão manipulados no projeto referente aos candidatos à vaga de Agente de Controle Urbano.

Este arquivo é uma lista contendo os campos:

> Inscrição (Número de sua inscrição no certame);

> Candidato (Nome do candidato);

> Nota Objetiva.

Em seguida o arquivo PDF (Resultado-preliminar-dos-Candidatos-Aprovados-OBJ) foi convertido para um TXT (Texto), para que assim os dados pudessem ser manipulados no Jupyter Notebook.

1. **TRATAMENTO DOS DADOS**

Nesta etapa realizamos a cópia dos arquivos contidos na etapa de preparação dos dados e em seguida utilizando do Jupyter Notebook, receberão o tratamento necessário para que posteriormente possam ser retirados os insights. Os dados serão catalogados conforme a data em que foram recebidos.

20180606

Foi recebido um arquivo TXT (Resultado-preliminar-dos-Candidatos-Aprovados-OBJ) do certame da SEDURB, contendo os dados que serão manipulados no projeto referente aos candidatos à vaga de Agente de Controle Urbano.

Este arquivo é uma lista contendo os campos:

> Inscrição (Número de sua inscrição no certame);

> Candidato (Nome do candidato);

> Nota Objetiva.

Em seguida o arquivo TXT (Resultado-preliminar-dos-Candidatos-Aprovados-OBJ) foi tratado com o Jupyter, que gerou um arquivo IPYNB (Tratamento - Lista de Aprovados SEDURBJP).

No arquivo IPYNB manipulamos os dados através de strings, que foram fatiadas diversas vezes até termos o resultado almejado, um Data Frame que continha as seguintes colunas:

> Número de Inscrição;

> Nota.

Por fim, foi gerado um CSV (Lista parcial de Aprovados na SEDURB) que continha os dados do Data Frame construído a partir da lista fornecida.

20180608

Realizada a cópia dos arquivos da última data, extraímos os dados do CSV gerado anteriormente (Lista parcial de Aprovados na SEDURB) através do arquivo IPYNB (Tratamento - Lista de Aprovados SEDURBJP).

Verificamos o Data Frame construído e ao o analisarmos percebemos que havia uma coluna desnecessária (Unnamed: 0) que pode ser excluída. Por fim salvamos novamente o arquivo.

20180611

Realizada a cópia dos arquivos da última data, extraímos os dados do CSV gerado anteriormente (Lista parcial de Aprovados na SEDURB) através do arquivo IPYNB (Tratamento - Lista de Aprovados SEDURBJP).

Verificamos o Data Frame construído e percebemos a necessidade de verificar se havia algum dado faltante no Data Frame, porém tudo estava perfeitamente preenchido. Por fim salvamos novamente o arquivo.

1. **INSIGHTS**

Nesta etapa realizamos a cópia dos arquivos contidos na etapa de tratamento dos dados e em seguida serão realizadas as análises para retirada dos insights. Os dados serão catalogados conforme a data em que foram recebidos.

20180606

Foi recebido um arquivo CSV (Lista parcial de Aprovados na SEDURB) do certame da SEDURB, contendo os dados que serão manipulados no projeto referente aos candidatos à vaga de Agente de Controle Urbano.

Este arquivo é um Data Frame contendo os campos:

> Número da Inscrição;

> Notas.

No arquivo IPYNB (Insights - Lista de Aprovados SEDURBJP) criado no Jupyter Notebook, foi verificado apenas que havia um campo desnecessário para realização dos insights.

20180608

Realizada a cópia dos arquivos da última data, extraímos os dados do CSV gerado anteriormente (Lista parcial de Aprovados na SEDURB) através do arquivo IPYNB (Insights - Lista de Aprovados SEDURBJP).

No Data Frame foi verificado uma coluna desnecessária, que em seguida foi excluída.

Passamos a realizar a retira dos insights os quais foram:

> A maior nota alcançada no certame e o(s) número(s) de inscrição relacionado(s):

> Número de inscrição: 351.966-0

> Nota: 89,00

> A menor nota alcançada no certame e o(s) número(s) de inscrição relacionado(s):

> Número de inscrição:

> 359.370-3

> 358.751-7

> 358.048-2

> 354.843-0

> 355.895-9

> 352.334-9

> 358.582-4

> 351.809-4

> 355.579-8

> 351.447-1

> 353.315-8

> 356.653-6

> 356.949-7

> 356.645-5

> 354.329-3

> 356.923-3

> 358.772-0

> 352.529-5

> 357.612-4

> 352.888-0

> 357.805-4

> 351.443-9

> 353.700-5

> 354.059-6

> 356.739-7

> 352.501-5

> 358.099-7

> 352.133-8

> 352.191-5

> 351.825-6

> 353.479-0

> 354.641-1

> 358.844-0

> 353.732-3

> 351.486-2

> 356.520-3

> Nota: 50,00

Como podemos observar, alguns candidatos tiraram a nota mínima para estar apto para a fase seguinte do exame.

20180609

Realizada a cópia dos arquivos da última data, extraímos os dados do CSV gerado anteriormente (Lista parcial de Aprovados na SEDURB) através do arquivo IPYNB (Insights - Lista de Aprovados SEDURBJP).

Por fim realizamos a analise principal do projeto, ordenamos os candidatos pelas notas e extraímos apenas os candidatos aptos para segunda fase, pois mesmo com uma nota superior a cinquenta pontos, os candidatos precisam estar classificados entre as 266 vagas. Em seguida salvamos o arquivo em um novo CSV (Lista dos candidatos aptos para segunda fase da SEDURB).

1. **FINAL**

Nesta etapa realizamos a cópia dos arquivos contidos na etapa de insights, este é o arquivo final do projeto, a conclusão do objetivo inicial do mesmo. Os dados serão catalogados conforme a data em que foram recebidos.

20180609

Foi recebido um arquivo CSV (Lista dos candidatos aptos para segunda fase da SEDURB) do certame da SEDURB, contendo os dados ordenados por nota dos 266 candidatos à vaga de Agente de Controle Urbano.

1. **CONCLUSÃO**

Ao elaborar o arquivo final (Lista dos candidatos aptos para segunda fase da SEDURB) concluímos o objetivo do projeto que consistia em uma lista ordenada pela nota dos candidatos, contendo apenas os candidatos aptos para próxima fase do certame.

Por se tratar de um projeto simples apenas algumas técnicas foram aplicadas, com ênfase na manipulação de strings.

Como ultima observação temos as lições aprendidas durante o projeto, das quais a mais importante é com certeza sobre o tratamento dos dados. Quando manipulamos strings percebemos o quão custoso foi esta forma de manipular os dados, mesmo sendo apenas uma lista com poucos campos a serem tratados. Verifica-se, portanto, a necessidade de adoção de novas ferramentas para realizar essa tarefa, uma das possibilidades é o Orange 3 que será analisado para uso nos próximos projetos.

* 1. Considerações do Autor

O Projeto SEDURB trate-se de um projeto piloto do meu portfólio. A estrutura adotada para o seu desenvolvimento tem como principal objetivo servir como modelo para os trabalhos posteriores deste mesmo portfólio.