da Indústria Indústria Indústria borracha, Indústria do do Minerais de da fumo, **Extrativa** Indústria Indústria material papel, **Bairros** nãomaterial madeira e couros, elétrico e de mineral metalúrgica mecânica papelão, metálicos de do peles, comunicações editorial transporte mobiliário similares e gráfica е diversas 0 NaN 2168 5022 12145 9180 4182 7394 2526 19329 11540 1 Total 2 NaN Área de Planejamento 1054 515 950 2167 627 915 157 7821 2238 Região de Planejamento 1054 515 950 2167 627 915 157 7821 2238 1.1 5 rows × 26 columns dataframe_2005.drop([0, 1, 2, 3, 4, 5, 10, 12, 17, 22, 24, 26, 27, 28, 29, 38, 41, 49, 51, 52, 56, 61, 62, 63, 64 166, 177, 179, 180, 189, 190, 191, 192, 197, 204, 205, 211, 212, 216, 217, 221], inplace=True) dataframe_2005 head() Indústria da Indústria Indústria Indústria borracha, Indústria do do **Minerais** de da fumo, Extrativa Indústria Indústria material papel, **Bairros** material nãomadeira e couros, mineral elétrico e de papelão, metalúrgica mecânica metálicos de do peles, comunicações editorial transporte mobiliário similares e gráfica е diversas 3 Saúde 9 11 13 12 Gamboa 8 30 8 103 Santo 83 47 48 132 8 39 157 Cristo 14 742 742 9 Caju 132 384 1172 133 161 34 4557 849 659 301 11 Centro 5 rows × 26 columns dataframe_2005.columns Index(['Bairros', 'Extrativa mineral', 'Minerais não-metálicos', 'Indústria metalúrgica', 'Indústria mecânica', 'Indústria do material elétrico e de comunicações', Indústria de material de transporte', 'Indústria da madeira e do mobiliário', 'Indústria do papel, papelão, editorial e gráfica', 'Indústria da borracha, fumo, couros, peles, similares e diversas', 'Indústria química de produtos farmacêuticos, veterinários, perfumaria...', 'Indústria têxtil do vestuário e artefatos de tecidos', 'Indústria de calçados', 'Indústria de produtos alimentícios, bebidas e álcool etílico', 'Serviços industriais de utilidade pública ', 'Construção civil', 'Comércio varejista', 'Comércio atacadista', 'Instituições de crédito, seguros e capitalização', 'Comércio e administração de imóveis, valores mobiliários, serviços técnicos', 'Transportes e comunicações', 'Serviços de alojamento, alimentação, reparação, manutenção, redação...', 'Serviços médicos, odontológicos e veterinários', 'Ensino', 'Administração pública direta e autárquica', 'Agricultura, silvicultura, criação de animais, extrativismo vegetal'], dtype='object') df_melted_2005 = pd.melt(dataframe_2005, id_vars =['Bairros'], value_vars =['Extrativa mineral', 'Minerais não-m ed_2005 = pd.mert(datarrame_2005, 10_vars = [thrives], varac_vars = [thrives], varse_vars = df_melted_2005.sort_values(by=['Bairros']) **Bairros** Atividade-Economica Numero-Empregos 2003 Abolição Indústria de produtos alimentícios, bebidas e ... 14 383 Abolição Indústria metalúrgica 17 Instituições de crédito, seguros e capitalização **2813** Abolição 88 Abolição Comércio varejista 1451 2489 **1031** Abolição Indústria da madeira e do mobiliário Água Santa Indústria de material de transporte 866 **3296** Água Santa Serviços de alojamento, alimentação, reparação... 9 Água Santa Minerais não-metálicos 218 **2000** Água Santa Indústria de produtos alimentícios, bebidas e ... **3134** Água Santa Transportes e comunicações 287 4050 rows × 3 columns df_melted_2005['Numero-Empregos'].replace('-', 0, inplace=True) file_path_bairros pd read_csv(file_path_bairros) bairros = bairros.head(3) AP **RP** RA Bairros Área de Planejamento 1 Região de Planejamento 1.1 I Portuária 1 Área de Planejamento 1 Região de Planejamento 1.1 I Portuária Gamboa Área de Planejamento 1 Região de Planejamento 1.1 I Portuária Santo Cristo complete_2005 = pd.merge(df_melted_2005, bairros, on='Bairros') complete_2005.head() Numero-**AP Bairros** Atividade-Economica **RP** RA **Empregos** Área de Planejamento Região de Planejamento Saúde Extrativa mineral Portuária Área de Planejamento Região de Planejamento Saúde Minerais não-metálicos Portuária 1.1 Região de Planejamento Área de Planejamento Indústria metalúrgica Saúde Portuária Região de Planejamento Área de Planejamento Saúde 9 Indústria mecânica Portuária Indústria do material elétrico e de Região de Planejamento Área de Planejamento Saúde 11 comunicações Portuária complete_2005[complete_2005 = complete_2005[['Ano', 'AP' complete_2005 **Numero-**Ano AP **RP** RA **Bairros** Atividade-Economica **Empregos** Região de Area de 0 2005 I Portuária Saúde Extrativa mineral 0 Planejamento 1 Planejamento 1.1 Área de Região de 1 2005 I Portuária Saúde Minerais não-metálicos Planejamento 1 Planejamento 1.1 Região de Area de I Portuária Indústria metalúrgica 2 2005 Saúde Planejamento 1 Planejamento 1.1 Região de Área de 9 3 2005 I Portuária Saúde Indústria mecânica Planejamento 1 Planejamento 1.1 Area de Região de Indústria do material elétrico e de 2005 11 4 I Portuária Saúde Planejamento 1 Planejamento 1.1 comunicações Região de **XXVI** Pedra de Serviços de alojamento, alimentação, Área de 120 3995 2005 Planejamento 5 Planejamento 5.4 Guaratiba Guaratiba reparação... Região de **XXVI** Serviços médicos, odontológicos e Area de Pedra de 3996 2005 6 Planejamento 5 Planejamento 5.4 veterinários Guaratiba Guaratiba Região de **XXVI** Pedra de Área de 2005 3997 Ensino 50 Planejamento 5 Planejamento 5.4 Guaratiba Guaratiba Região de **XXVI** Pedra de Administração pública direta e Area de 3998 2005 0 Planejamento 5 Planejamento 5.4 Guaratiba Guaratiba autárquica **XXVI** Agricultura, silvicultura, criação de Área de Região de Pedra de 3999 2005 Planejamento 5 Planejamento 5.4 Guaratiba Guaratiba animais,... 4000 rows × 7 columns empregados = pd.DataFrame() anos = ['2005', '2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015', '2016', '2017', '2018'] for ano in anos: df = pd.read_excel(file_path_empregados, sheet_name=ano, header=3, nrows=224, usecols="A:Z") df.drop([0, 1, 2, 3, 4, 5, 10, 12, 17, 22, 24, 26, 27, 28, 29, 38, 41, 49, 51, 52, 56, 61, 62, 63, 64, 166, 177, 179, 180, 189, 190, 191, 192, 197, 204, 205, 211, 212, 216, 217, 221], inplace=True) df_melt = pd.melt(df, id_vars =['Bairros'], value_vars =['Extrativa mineral', 'Minerais não-metálicos',
 'Indústria metalúrgica', 'Indústria mecânica', 'Indústria química de produtos farmacêuticos, veterinários, perfumaria...',
 'Indústria têxtil do vestuário e artefatos de tecidos',
 'Indústria de calçados',
 'Indústria de produtos alimentícios, bebidas e álcool etílico',
 'Serviços industriais de utilidade pública ', 'Construção civil',
 'Comércio varejista', 'Comércio atacadista',
 'Instituições de crédito, seguros e capitalização',
 'Comércio e administração de imóveis, valores mobiliários, serviços técnicos',
 'Transportes e comunicações',
 'Serviços de alojamento, alimentação, reparação, manutenção, redação...',
 'Serviços médicos, odontológicos e veterinários', 'Ensino',
 'Administração pública direta e autárquica',
 'Agricultura, silvicultura, criação de animais, extrativismo vegetal'],
 var_name ='Atividade-Economica', value_name = "Numero-Empregos")

df_melt['Numero-Empregos'] replace('-', 0, inplace=True)

df_merged = pd.merge(df_melt, bairros, on='Bairros') df_merged = pd.merge(df_melt, bairros, on='Bairros') df_merged['Ano'] = ano df_merged = df_merged[['Ano', 'AP', 'RP', 'RA', 'Bairros', 'Atividade-Economica', 'Numero-Empregos']] empregados = empregados.append(df_merged,ignore_index=True) empregados **Numero-**Atividade-Economica Ano AP RP RA **Bairros Empregos** Área de Região de I Portuária Extrativa mineral 0 2005 Saúde 0 Planejamento 1.1 Planejamento 1 Área de Região de 2005 I Portuária Minerais não-metálicos 5 Saúde 1 Planejamento 1 Planejamento 1.1 Área de Região de 2 2 2005 I Portuária Saúde Indústria metalúrgica Planejamento 1.1 Planejamento 1 Região de Área de 2005 9 3 Saúde Indústria mecânica I Portuária Planejamento 1 Planejamento 1.1 Área de Indústria do material elétrico e de Região de 2005 I Portuária Saúde 11 4 Planejamento 1.1 comunicações Planejamento 1

XXVI

XXVI

XXVI

IVXX

IVXX

Guaratiba

Guaratiba

Guaratiba

Guaratiba

Guaratiba

Região de

Região de

Região de

Região de

Região de

Planejamento 5.4

Planejamento 5.4

Planejamento 5.4

Planejamento 5.4

Planejamento 5.4

Área de

Área de

Área de

Área de

Área de

Planejamento 5

Planejamento 5

Planejamento 5

Planejamento 5

Planejamento 5

2018

2018

2018

2018

56000 rows × 7 columns

lean.csv', index = False)

55995

55996

55998

55999

55997 2018

Pedra de

Guaratiba

empregados.to_csv(r'/home/blackmamba/data_analysis/atividades_econ_rio/clean_data/atividades_econ_ri_ap_bairros_c

Serviços de alojamento, alimentação,

Serviços médicos, odontológicos e

Administração pública direta e

Agricultura, silvicultura, criação de

reparação...

veterinários

autárquica

animais,...

Ensino

118

12

201

0

0

file_path_empregados = "raw_data/atividades_econ_rj_ap.xls"

dataframe_2005 head()

dataframe_2005 = pd.read_excel(file_path_empregados, sheet_name="2005", header=3, nrows=224, usecols="A:Z")

Indústria