

Distribuições

```
import argparse
import logging
import os
from utils import load_csv
import seaborn as sns
import matplotlib.pyplot as plt

def setup_logging(verbose: bool):
    os.makedirs("../logs", exist_ok=True)
    logging.basicConfig(
        level=logging.DEBUG if verbose else logging.INFO,
        format="%(asctime)s - %(levelname)s - %(message)s",
        handlers=[
            logging.FileHandler("logs/execucao.log", mode='w'),
            logging.StreamHandler()
        ]
    )

def parse_args():
    parser = argparse.ArgumentParser(description="Script para análise de CSVs")

    parser.add_argument(
        "--input",
        type=str,
        required=True,
        help="Caminho para o arquivo CSV de entrada."
    )
    parser.add_argument(
        "--verbose",
        action="store_true",
        help="Se definido, mostrará logs de debug mais detalhados."
    )
```

```

    return parser.parse_args()

def main():

    logging.info("Início da execução")
    path_file = r'../data/raw_data/census.csv'
    df = load_csv(path_file)
    logging.info(f"DataFrame carregado com {df.shape[0]} linhas e {df.shape[1]} colunas.")
    logging.info(f"Primeiras 5 linhas:\n{df.head()}")
    # Verificar o tipo de distribuição para cada um dos atributos do DataFrame, usando seaborn

    for column in df.select_dtypes(include=['float64', 'int64']).columns:
        logging.info(f"Analisando a coluna: {column}")
        sns.histplot(df[column], kde=True)
        plt.title(f"Distribuição da coluna {column}")
        plt.xlabel(column)
        plt.ylabel("Frequência")
        #plt.savefig(f"logs/distribuicao_{column}.png")
        plt.show()
        plt.close()
        logging.info(f"Gráfico salvo como logs/distribuicao_{column}.png")

    logging.info("Fim da execução")

```