CODEBOOK FOR DATA ANALYSIS

DICTIONARIES

TRANSACTION AND LOGIN DATA

Our raw data is organized into three main folders: “output--transactions_20160826”, “output--valuations_holdings”, and “output--logins”. Each folder contains separate files for every investor.

Files in “output--transactions_20160826”

Overview: The folder contains individual files for each investor. These files provide detailed information about every transaction made by the investor.

Column Descriptions

- **Date**: Represents the date when the transaction took place.
- **transaction_time**: The exact time when the transaction was recorded. It is presented in a 24-hour clock format: HHMM.
- **transaction_type**: This indicates the nature of the transaction, specifying whether it was a purchase or a sale. The values include ‘B’ for Buy and ‘S’ for Sell.
- **narrative**: This column contains descriptive text related to the transaction. Examples include ‘BUY’ for the purchase of securities, ‘SELL’ for the sale of securities, ‘MOVEMENT ON’ for other types of securities movements or adjustments, and ‘CONSOL’ for the consolidation of holdings.
- **sedol**: An abbreviation for Stock Exchange Daily Official List, this is a code that represents the security that was traded in the transaction.
- **cost**: This shows the total cost of the transaction, denominated in pounds.
- **quantity**: The number of units that were bought or sold in the transaction.
- **unit_price**: This refers to the price per unit of the security at the time of the transaction.
- **commission**: The commission fee that is associated with the transaction.

Files in “output--valuations_holdings”

Overview: The folder contains individual files for each investor. These files provide detailed information about the portfolios held by each investor on specific days, which may or may not coincide with transaction dates.

Column Descriptions

- **ValueDate**: The date on which the value of the asset was recorded.
- **Sedol**: An acronym for Stock Exchange Daily Official List, it is a code that uniquely identifies a security.
- **Isin**: International Securities Identification Number, which also uniquely identifies a security.
- **asset_class_name**: The classification of the asset. For example, this could be ‘EQUITY’ indicating stocks.
- **Quantity**: The number of units of the asset held.

- **Bookcost**: The original cost of acquiring the asset.
- **last_trade_date**: The date of the last trade for the asset.
- **Price**: The price of the asset per unit at the last trade date.
- **CalculatedValue**: The total value of the holding, calculated as the Quantity multiplied by the Price.

Files in "output--logins"

Overview: The folder contains individual files for each investor, detailing their login activity for each day they logged in.

Column Descriptions

- **d**: This column represents the date of login activity.
- **nlogins**: The number of logins recorded on the specified date.

DEMOGRAPHICS

File: 'anonymous_customer_level_data_merged_master.csv'

Overview: The file holds demographic information for all the accounts in the portfolio data.

Column Descriptions

- **anon_portfolio_id**: A unique identifier assigned to each customer's portfolio.
- **anon_customer_id**: A unique identifier allocated to each customer.
- **gender**: The gender of the customer. The possible values are 'M' for Male and 'F' for Female.
- **postcode**: The postal code corresponding to the customer's registered address.
- **dob**: The year of birth of the customer. It is recorded in the format YYYY.
- **All_Holdings_Value_Amount**: This represents the total value of all holdings within the customer's portfolio.
- **status**: Indicates the current status of the customer's account. Possible values include 'complete', 'incomplete', and 'screened'.
- **portfolio_open_date**: The date when the customer's portfolio was initially opened.
- **portfolio_type**: The type of portfolio held by the customer. Options include 'MarketMaster', 'Retail ISA', 'SIPP', and others.

File: 'anonymous_customer_level_data_merged_master_login.csv'

Overview: This file contains demographic information associated with all accounts included in the portfolio data, specifically for those accounts where login records are available. The columns in this file are similar to those in the previously mentioned demographic data file.

PRICE DATA

The following files contain detailed stock data sourced from Datastream.

File: 'all_sedols_isin_matching.csv'

Overview: This file is a matching table between ISIN (International Securities Identification Number) and SEDOL (Stock Exchange Daily Official List) codes.

Column Descriptions

- **sedol**: This column contains the Stock Exchange Daily Official List (SEDOL) number.
- **isin**: This column features the International Securities Identification Number (ISIN).
- **num_sedols_per_isin**: This represents the number of SEDOLs that are associated with a single ISIN. It indicates the presence of multiple SEDOL codes for a single ISIN.
- **num_isins_per_sedol**: This shows the number of ISINs that are associated with a single SEDOL. It indicates the presence of multiple ISIN codes for a single SEDOL.

File: 'datastream_20180601.csv'

Overview: This file contains price data for each stock.

Column Descriptions:

- **Code_used_DS:** This is the unique Datastream identifier for the security used within the dataset.
- **date:** The date associated with the price data.
- **af:** Stands for the Adjustment Factor. It is a numerical value used to adjust historical financial data for a stock. This adjustment accounts for corporate actions such as stock splits, dividends, rights issues, and consolidations.
- **upp:** Unadjusted price, not adjusted for dividends or splits.
- **app:** Adjusted price, adjusted for dividends or splits.
- **DS_STOCK_TYPE:** Indicates the type of stock for the security. Examples include 'EQ' for Equity, 'UT' for Unit Trust, 'BD' for Bond, 'EWT' for Equity Warrant, and 'INVT' for Investment Trust.
- **DS_STOCK_TYPE_DESC:** This shows the description of the stock type, like 'Equity' for 'EQ', 'Unit Trust' for 'UT', etc.
- **DS_SECURITY_TYPE_CODE:** A code representing the security type. For example, 'ORD' suggests an ordinary share. Other examples include 'TRAD' for Traditional Warrants, 'ETFE' for Equity Exchange-Traded Fund, 'CEF' for Closed-Ended Fund, etc.

- **DS_SECURITY_TYPE:** Describes the type of security, such as ‘Ordinary Shares’, ‘Traditional Warrants’, etc.
- **DS_GEOGRAPHY_GROUP_NAME:** Indicates the geographical market or region where the security is listed or primarily traded. Examples include ‘United Kingdom’, ‘Ireland’, ‘Luxembourg’, ‘France’, ‘United States’, etc.
- **sedol:** The Stock Exchange Daily Official List (SEDOL) code for the security.

File: ‘ftse100 return.csv’

Overview: This file contains historical data related to the FTSE 100 index.

Column Descriptions:

1. **valuedate_c**: The date for each data entry.
2. **return_FTSE100**: The return of the FTSE 100 index for each respective date. The return is expressed as a decimal, representing the percentage change in the index value from the previous day.

File: ‘260118 isins sedols recovered static and industry.csv’

Overview: The file includes detailed classifications into various industry groups and subgroups

Column Descriptions

- **code_used_ds**: Unique identifier used within Datastream for the security.
- **FTSECTOR**: FTSE sector classification.
- **INDUSTRYGROUP**: Industry group classification.
- **ICBCODE**: International Classification Benchmark code.
- **ICBINDUSTRYCODE**: ICB industry code.
- **ICBINDUSTRYNAME**: Name of the industry according to the ICB classification.
- **TRBCECONSECCODE**: TRBC (Thomson Reuters Business Classification) economic sector code.
- **TRBCECONSECNAME**: Name of the economic sector according to TRBC.
- **INDUSTRY_GROUP2_NEW**: Secondary industry group classification, providing more specific categorization. Examples include: Exchange Traded Funds, Investment Trusts, Exchange Traded Commodity, Specialty Finance, Business Support Services, Exploration & Production, etc.