



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

Pós-graduação *Lato Sensu* em Analytics e Business Intelligence

RELATÓRIO TÉCNICO

ANÁLISE DE DADOS DE FUNDOS IMOBILIÁRIOS
NA BOLSA DE VALORES BRASILEIRA

LUIZ AUGUSTO ZELAQUETT DE SOUZA

Belo Horizonte

2022

SUMÁRIO

1. Introdução	3
1.1. Contexto	3
1.2. Objetivos	3
1.3. Público alvo	3
2. Modelo de Dados	4
2.1. Modelo Dimensional	4
2.2. Fatos e Dimensões	4
3. Integração, Tratamento e Carga de Dados	8
3.1. Fontes de Dados.....	8
3.2. Processos de Integração e Carga (ETL).....	8
3.2.1 Etapa 1	9
3.2.2 Etapa 2	10
4. Camada de Apresentação	10
4.1 Dashboards	10
4.1.1 Dashboard Estratégico	11
4.1.2 Dashboard Tático.....	12
4.1.2 Dashboard Operacional	13
4.2 Análises avançadas	14
5. Registros de Homologação	22
5. Conclusões	24
6. Trabalhos Futuros	25
7. Links	26
REFERÊNCIAS	27

1. Introdução

1.1. Contexto

O presente projeto visa apresentar dados relativos à característica e a negociação e de Fundos Imobiliários – FIs na Bolsa de valores brasileira (B3) e que possam ser utilizados como escolha ou como auxílio de futuros investimentos financeiros.

Utilizando dados públicos, disponibilizados em <http://dados.cvm.gov.br/dataset/> pela Comissão de Valores Mobiliários – CVM, em <https://www.fundsexplorer.com.br> pelo site referência no mercado de FIs no Brasil e no Kaggle.

1.2. Objetivos

Este projeto tem como objetivo apresentar a pessoas interessadas em realizar investimentos em Fundos Imobiliários as características mais importantes desses ativos financeiros, assim como definir quais desses fundos, no presente momento, são os mais relevantes a serem analisados para futuros investimentos financeiros.

1.3. Público alvo

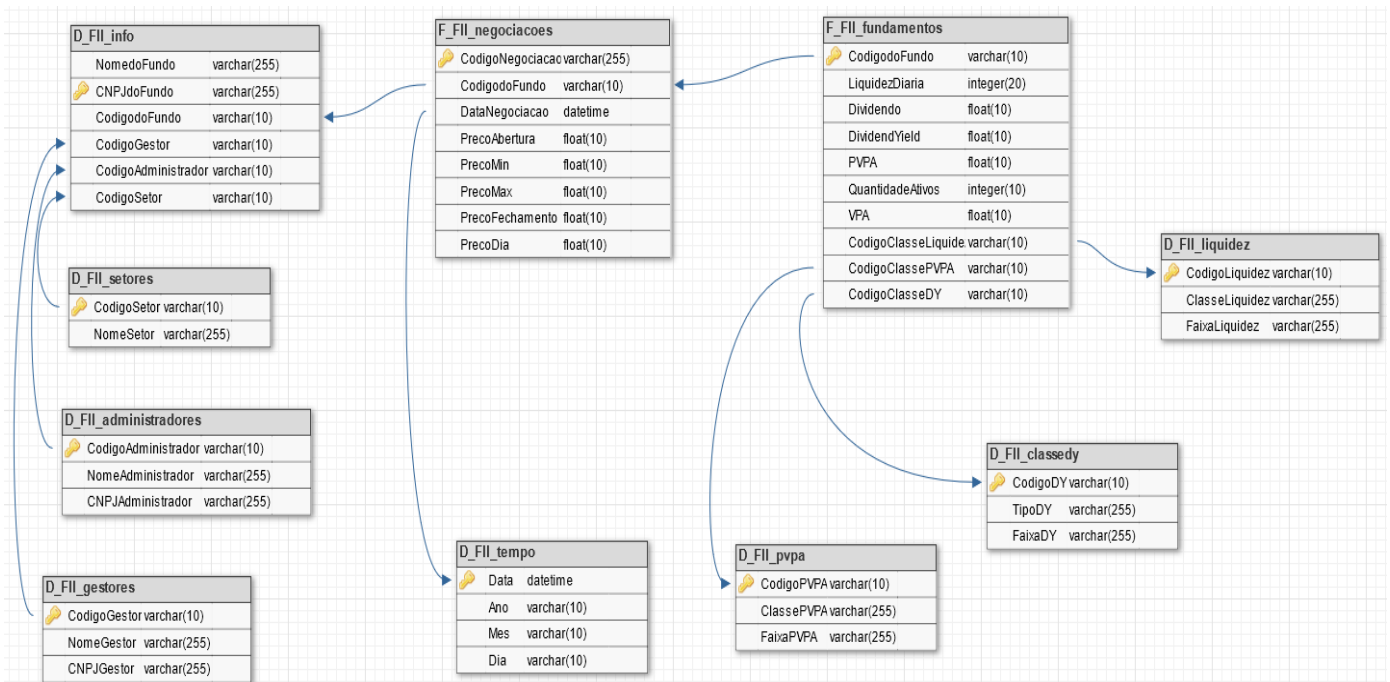
O público alvo são as pessoas interessadas em realizar investimentos em Fundos Imobiliários.

O perfil de conhecimento dessas pessoas deve ser intermediário, pois existem vocabulários específicos referentes a aplicações financeiras, bem como sempre deve ser ressaltado que Fundos Imobiliários são ativos de renda variável, negociados na bolsa de valores e que nunca rentabilidade ou característica passadas ou atuais do Fundos Imobiliários, são garantias de rentabilidade ou características futuras.

2. Modelo de Dados

2.1. Modelo Dimensional

Etapa de grande importância para os projetos de *Business Intelligence*, a modelagem dimensional tem como objetivo estruturar as entidades das bases de dados, onde o mesmo possui comportamento diferente de tabelas fato e dimensão.



2.2. Fatos e Dimensões

As tabelas mostradas no diagrama se dividem em fatos e dimensões, sendo elas descritas com respectivas descrições dos seus conteúdos, abaixo:

D_FII_info:

Dimensão que contém informações intrínsecas ao Fundo Imobiliário.

Nome do Campo	Tipo	Descrição
NomedoFundo	Varchar(255)	Nome do Fundo junto à CVM
CNPJdoFundo	Varchar(255)	CNPJ do Fundo junto à CVM
CodigodoGestor	Varchar(10)	Código Interno
CodigoAdministrador	Varchar(10)	Código Interno
CodigoSetor	Varchar(10)	Código Interno

D_FII_setores:

Dimensão que contém as informações relativas ao setor ao qual o do Fundo Imobiliário pertence.

Nome do Campo	Tipo	Descrição
CodigoSetor	Varchar(10)	Nome do Fundo junto à CVM
NomeSetor	Varchar(255)	Setor ao qual o fundo imobiliário pertence

D_FII_administrador:

Dimensão que contém as informações relativas ao administrador do Fundo Imobiliário

Nome do Campo	Tipo	Descrição
CodigoAdministrador	Varchar(10)	Código interno
NomeAdministrador	Varchar(255)	Nome do administrador Fundo junto à CVM
CNPJAdministrador	Varchar(255)	CNPJ do administrador do Fundo junto à CVM

D_FII_gestor:

Dimensão que contém as informações relativas ao gestor do Fundo Imobiliário

Nome do Campo	Tipo	Descrição
CodigoGestor	Varchar(10)	Código interno
NomeGestor	Varchar(255)	Nome do gestor do Fundo junto à CVM
CNPJGestor	Varchar(255)	CNPJ do gestor do Fundo junto à CVM

F_FII_negociacoes:

Fato que contém as informações relativas à negociação do Fundo Imobiliário na bolsa de valores.

Nome do Campo	Tipo	Descrição
CodigoNegociacao	Varchar(255)	Código interno
CodigodoFundo	Varchar(10)	Identificador do fundo no padrão XXXX11
DataNegociacao	Datetime	Data da negociação
PrecoAbertura	Float(10)	Preço de abertura do fundo
PrecoMin	Float(10)	Preço mínimo do fundo
PrecoMax	Float(10)	Preço máximo do fundo
PrecoFechamento	Float(10)	Preço de fechamento do fundo

F_FII_fundamentos:

Fato que contém as informações relativas aos fundamentos do Fundo Imobiliário.

Nome do Campo	Tipo	Descrição
CodigodoFundo	Varchar(10)	Identificador do fundo no padrão XXXX11
LiquidezDiaria	Integer(20)	O quanto em valor é negociado do fundo na bolsa
Dividendo	Float(10)	Último valor pago de dividendo
DividendYield	Float(10)	DividendYield atual do fundo
PVPA	Float(10)	Preço sobre valor patrimonial
QuantidadeAtivos	Integer(10)	Quantidade de ativos pertencentes ao fundo
VPA	Float(10)	Valor patrimonial do fundo
CodigoClasseLiquidez	Varchar(10)	Código Interno
CodigoClassePVPA	Varchar(10)	Código Interno
CodigoClasseDY	Varchar(10)	Código Interno

D_FII_data:

Dimensão de tempo

Nome do Campo	Tipo	Descrição
Data	Datetime	
Dia	Integer(2)	
Mes	Integer(2)	
Ano	Integer(4)	

D_FII_pvpa:

Dimensão que classifica o PVPA (Preço de negociação sobre Valor Patrimonial) do fundo em baixa, média e alta. Dependendo de um range entre os PVPAs dos fundos negociados na bolsa de valores.

Nome do Campo	Tipo	Descrição
CodigoPVPA	Varchar(10)	Código interno
ClassePVPA	Varchar(255)	Tipos de classes de PVPA: baixa, média e alta.
FaixaPVPA	Varchar(255)	Determinação do intervalo de cada uma das classes.

D_FII_classeDY:

Dimensão que classifica o Dividend Yield do fundo em baixa, média e alta. Dependendo de um range entre os DYs dos fundos negociados na bolsa de valores.

Nome do Campo	Tipo	Descrição
CodigoDY	Varchar(10)	Código interno
TipoDY	Varchar(255)	Tipos de classes de dy: baixo, médio e alto.
FaixaDY	Varchar(255)	Determinação do intervalo de cada uma das classes.

D_FII_liquidez:

Dimensão que classifica a liquidez do fundo em baixa, média e alta. Dependendo de um range de negociações efetuadas na bolsa.

Nome do Campo	Tipo	Descrição
CodigoLiquidez	Varchar(10)	Código interno
ClasseLiquidez	Varchar(255)	Tipos de classes de liquidez: baixa, média e alta.
FaixaLiquidez	Varchar(255)	Determinação do intervalo de cada uma das classes.

3. Integração, Tratamento e Carga de Dados

3.1. Fontes de Dados

Para esta análise, os dados serão extraídos de três fontes, sendo:

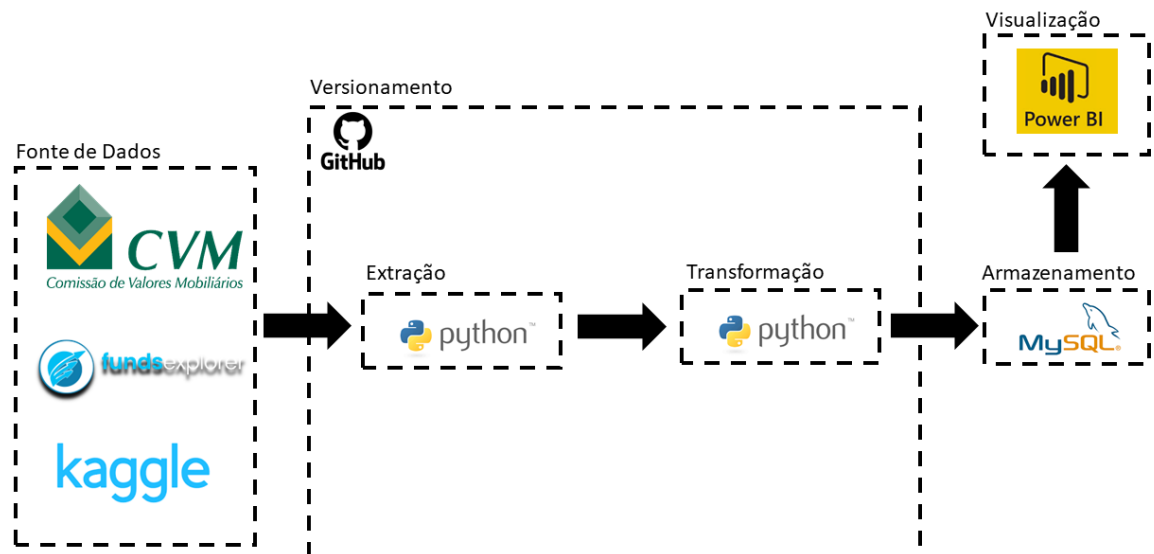
- 1- Dados contidos no Kaggle, em <https://www.kaggle.com/maverickjpa/fundos-imobiliarios-brasil-brazilian-funds-fiis> ;
- 2- Dados da CVM contidos em: <http://dados.cvm.gov.br/dataset/> ;
- 3- Dados contidos no site www.fundsexplorer.com.br, que têm como finalidade a construção dos ranges dos dy, pvpa e liquidez.

3.2. Processos de Integração e Carga (ETL)

O processo de ETL foi realizado em Python com a utilização das bibliotecas panda e numpy. Foram concatenados os arquivos referentes as negociações aos fundamentos e as informações dos Fundos Imobiliários e, na sequência criadas as dimensões.

Cabe destacar que foram criados códigos para as colunas D_FII_pvpa, D_FII_classeDY e D_FII_liquidez uma vez que não existiam na base original.

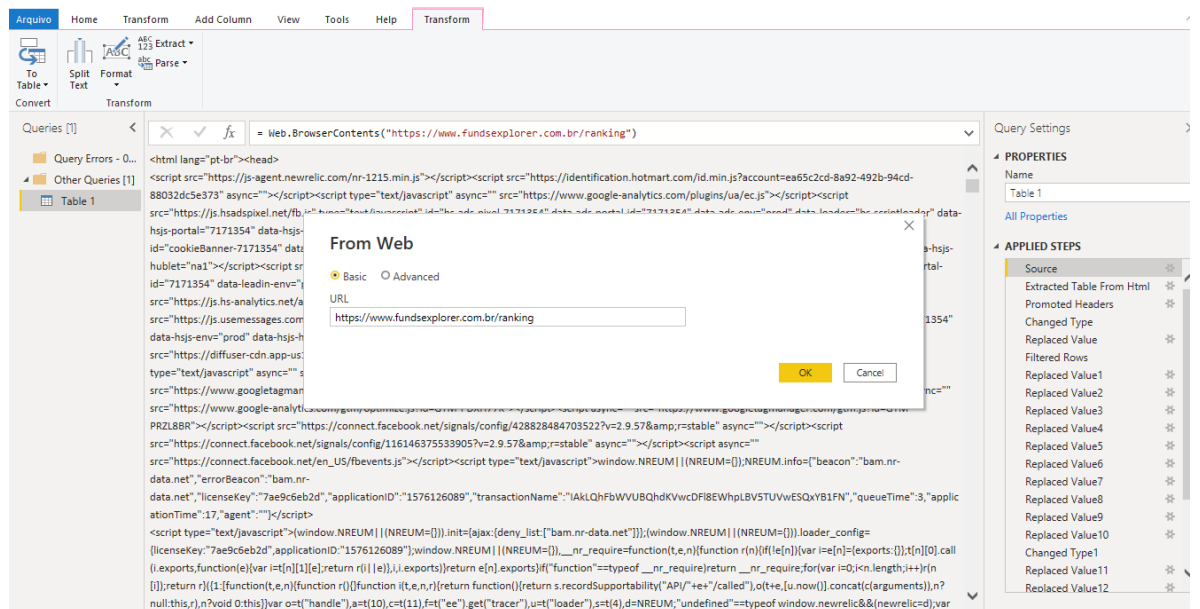
Abaixo segue como foi arquitetado o fluxo de ETL do projeto:



3.2.1 Etapa 1

Foi realizada a importação seguindo o processo abaixo:

- 1- Obter dados da web;
- 2- Inserção do site com a fonte dos dados utilizados.



3.2.2 Etapa 2

Após o carregamento da fonte de dados ter sido realizado, foram realizadas algumas transformações nos dados carregados:

- 1- Substituição de todos valores “N/A” e NULOS, de campos passíveis (DY, DY(12), DY(3M), DU(6M), P/VPA e Dividendo) de cálculo, por 0 para que os cálculos, como os das médias, fossem realizados sem erro;
- 2- Substituição de valores de setores NULOS, como sendo “outros”;
- 3- Os campos de porcentagem e que representavam moeda, foi necessário a retirada dos caracteres “%” e “R\$” para facilitar os cálculos;
- 4- A mudança de tipo foi necessária também nos campos mostrados no item 1, pois eles o PowerBI, no carregamento entendeu eles como sendo text.

O Resultado final do conjunto de dados utilizado foi o seguinte:

	Código fundo	Setor	Preço Atual	Liquidez Diária	Dividendo	Dividend Yield
1	ABCP11	Shoppings	RS 73,25	713.0		0,50
2	AFHI11	Títulos e Val. Mob.	RS 101,02	10478.0		1,30
3	AFOF11	Títulos e Val. Mob.	RS 94,58	207.0		1,00
4	AIEC11	Lajes Corporativas	RS 80,88	6115.0		1,11
5	ALMI11	Lajes Corporativas	RS 950,99	1.0		0,00
6	ALZR11	Logística	RS 115,50	9397.0		0,76
7	APTO11	Híbrido	RS 10,80	285.0		0,11
8	ARCT11	Híbrido	RS 106,43	22355.0		1,65
9	ARRI11	Títulos e Val. Mob.	RS 98,00	4875.0		1,34
10	ATSA11	Shoppings	RS 92,88	3.0		0,25
11	BAR11	Títulos e Val. Mob.	RS 100,85	31070.0		1,10
12	BBFI11B	Lajes Corporativas	RS 2.030,05	214.0		32,05
13	BBFO11	Títulos e Val. Mob.	RS 72,00	1169.0		0,72
14	BBIM11	Títulos e Val. Mob.	RS 0,0	0		0,25
15	BBPO11	Lajes Corporativas	RS 87,81	21599.0		0,89
16	BBRC11	Outros	RS 97,10	815.0		0,94
17	BCFF11	Títulos e Val. Mob.	RS 68,95	39394.0		0,54
18	BCIA11	Títulos e Val. Mob.	RS 81,46	4873.0		0,68
19	BCRI11	Títulos e Val. Mob.	RS 107,79	20568.0		1,35
20	BICE11	Títulos e Val. Mob.	RS 0,0	0		5,40
21	BICR11	Títulos e Val. Mob.	RS 99,49	118.0		0,68
22	BIME11	Outros	RS 9,32	25971.0		0,11
23	BLCA11	Lajes Corporativas	RS 97,99	10051.0		0,51

4. Camada de Apresentação

4.1 Dashboards

A ferramenta PowerBI foi utilizada para realizar o carregamento e tratamento dos dados, bem como o desenvolvimento, visualização e publicação dos dashboards.

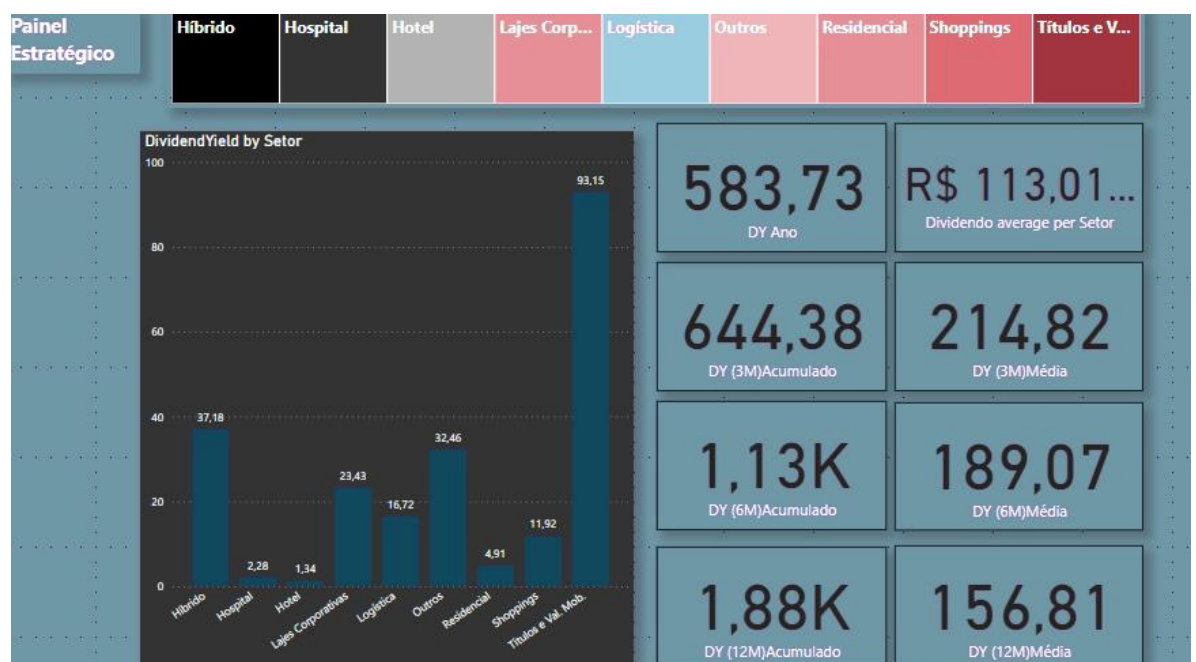
Os dashboards foram organizados nas seguintes categorias: estratégico, tático e operacional.

4.1.1 Dashboard Estratégico

Contendo uma visão mais macro do cenário dos Fundos Imobiliários, através de indicadores diversos, KPIs e principais visualizações dados consolidados para o contexto analisado.

No *dashboard* abaixo, é possível ter acesso as seguintes informações:

- Filtro por setor;
- Qual o Dividend Yield (DY) no ano de cada setor;
- Qual o dividendo médio em Real de cada setor;
- Qual o Dividend Yield acumulado nos últimos nos últimos 3 meses de cada setor;
- Qual o Dividend Yield médio nos últimos nos últimos 3 meses de cada setor;
- Qual o Dividend Yield acumulado nos últimos nos últimos 6 meses de cada setor;
- Qual o Dividend Yield médio nos últimos nos últimos 6 meses de cada setor;
- Qual o Dividend Yield acumulado nos últimos nos últimos 12 meses de cada setor;
- Qual o Dividend Yield médio nos últimos nos últimos 12 meses de cada setor.



4.1.2 Dashboard Tático

Contendo uma visão de acompanhamento histórico e segmentado das informações apresentadas no Dashboard Estratégico;

No dashboard abaixo, é possível ter acesso as seguintes informações:

- Filtro por setor;
- Filtro para cada fundo imobiliário, definido por seu código, pertencente ao setor definido;
- O Dividendo médio pago entre os fundos imobiliários, do setor selecionado, ou para apenas um subconjunto dos códigos selecionados;
- O preço unitário no mercado do fundo imobiliário ou a soma desses preços;
- O Dividendo ATUAL médio pago entre os Fundos imobiliários de um setor escolhido;
- O Dividendo ATUAL máximo pago entre os Fundos imobiliários de um setor escolhido;
- O Dividendo ATUAL mínimo pago entre os Fundos imobiliários de um setor escolhido;
- O P/VPA ATUAL médio entre os Fundos imobiliários de um setor escolhido;
- O P/VPA ATUAL máximo entre os Fundos imobiliários de um setor escolhido;
- O P/VPA ATUAL mínimo entre os Fundos imobiliários de um setor escolhido.



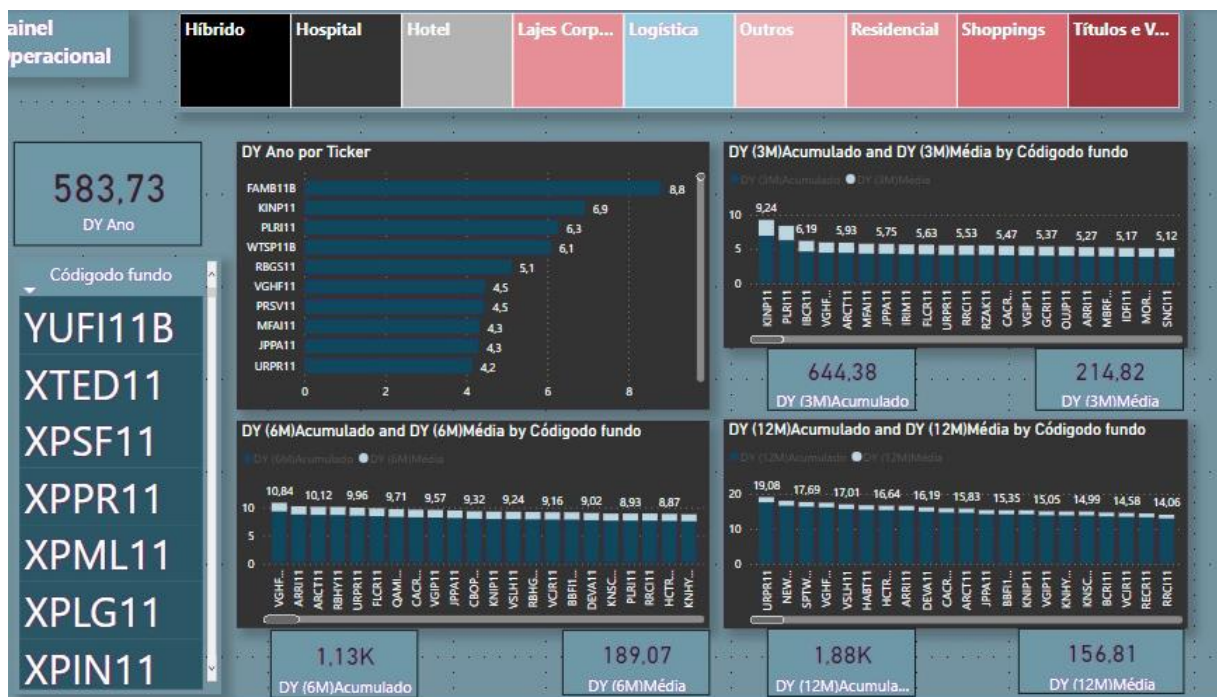
4.1.2 Dashboard Operacional

Painel Operacional - detalhamento dos dados apresentados nos painéis estratégico e tático.

No dashboard abaixo, é possível ter acesso as seguintes informações:

- Filtro por setor;
- Filtro para cada fundo imobiliário, definido por seu código, pertencente ao setor definido;
- O DY no ano para um setor ou grupo de Fundos Imobiliários selecionados;
- Uma visão gráfica com o DY no ano para um setor ou grupo de Fundos Imobiliários selecionados;
- Uma visão gráfica com o DY acumulado e médio nos últimos 3 meses, para um setor ou grupo de Fundos Imobiliários selecionados;

- O valor do DY acumulado e médio nos últimos 3 meses, para um setor ou grupo de Fundos Imobiliários selecionados;
- Uma visão gráfica com o DY acumulado e médio nos últimos 6 meses, para um setor ou grupo de Fundos Imobiliários selecionados;
- O valor do DY acumulado e médio nos últimos 6 meses, para um setor ou grupo de Fundos Imobiliários selecionados;
- Uma visão gráfica com o DY acumulado e médio nos últimos 12 meses, para um setor ou grupo de Fundos Imobiliários selecionados;
- O valor do DY acumulado e médio nos últimos 12 meses, para um setor ou grupo de Fundos Imobiliários selecionados.



4.2 Análises avançadas

As análises avançadas foram realizadas utilizando a linguagem Python no ambiente Colab do Google, e se subdividiu em duas etapas.

A primeira etapa consistiu em uma Análise Exploratória de Dados dos fundos imobiliários listados na bolsa de valores brasileira, a B3.

A análise está dividida em duas etapas:

1. Análise de uma base contendo os dados de FII's (Fundos de Investimento Imobiliários) em 2022 e em 2019;
2. Comparação entre as FII's de 2022 vs 2019.

Para as duas bases de dados, foram utilizadas as mesmas técnicas e representações para a visualização dos dados, de forma que seja estabelecida um padrão analítico único.

As análises realizadas foram:

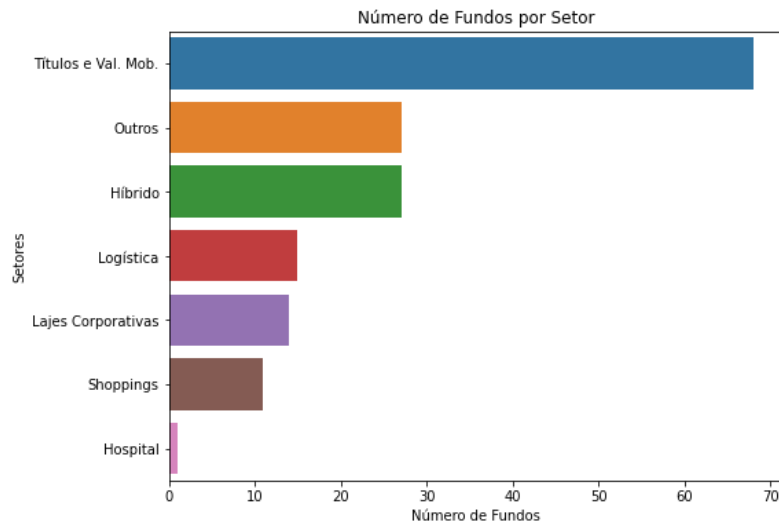
- Análise de dados categóricos;
- Análise de uma variável;
- Análise multivariável;
- Análise comparativa e variâncias.

Essas análises foram realizadas sobre as seguintes variáveis:

Variável	Tipo
Setor	Categórica
Preço Atual	Numérica
Liquidez Diária	Numérica
Dividendo	Numérica
DividendYield	Numérica
DY (3M) Média	Numérica
DY (6M) Média	Numérica
DY (12M) Média	Numérica
DY Ano	Numérica
Variação Preço	Numérica
Rentab.Período	Numérica
Rentab.Acumulada	Numérica
PatrimônioLiq.	Numérica
VPA	Numérica
P/VPA	Numérica
VacânciaFísica	Numérica
QuantidadeAtivos	Numérica

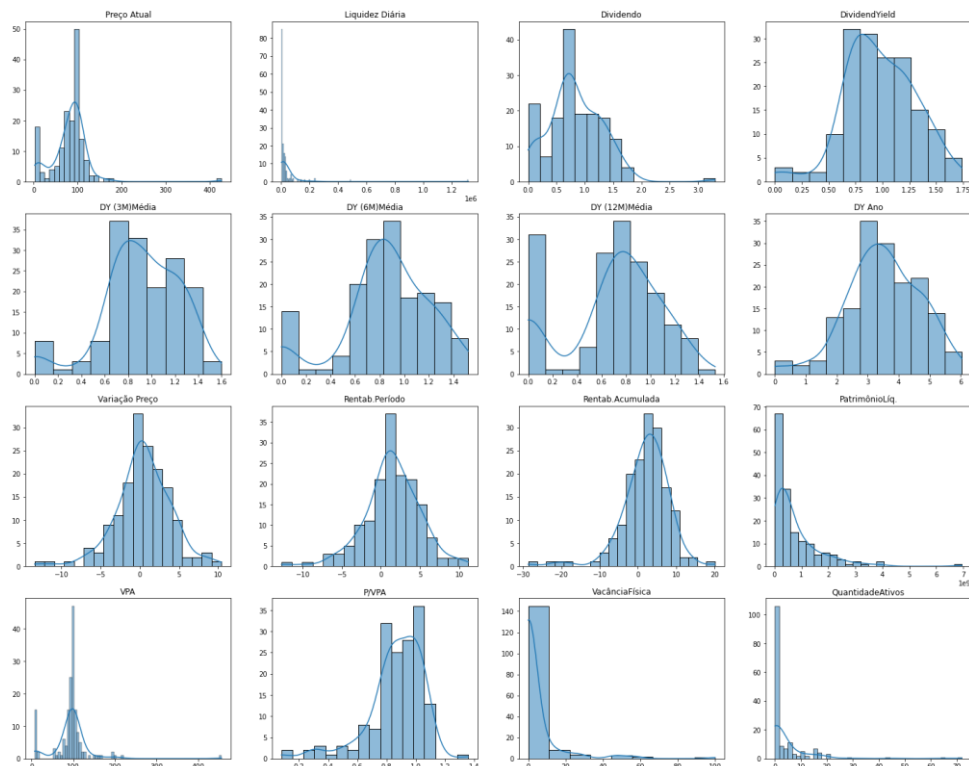
Na Análise de dados categóricos, foi analisada a distribuição dos FIIs pelos setores existentes, no caso, Títulos e Valores Mobiliários (Papel), Outros, Híbrido, Logístico, Lages Cooperativas, Shoppings e Hospitais.

Nesta análise pode-se ver uma concentração muito grande no setor de , Títulos e Valores Mobiliários.



Na etapa de análise das variáveis numéricas, primeiro foi analisado (por meio de boxplots, violinplots e histogramas) as variáveis individualmente.

Abaixo pode-se ver a distribuição das variáveis numéricas, utilizando histogramas.



O P/VPA, observa-se que a maioria dos FIs estão sendo negociados, hoje, abaixo do seu valor patrimonial ou muito próximo do valor patrimonial, o que é um indicativo de que os fundos estão baratos.

Uma análise também interessante é o confronto do DY ano com DY 12 meses. Essa análise mostra um aumento entre de DY entre todos os fundos, CERTAMENTE decorrente da melhoria de toda conjuntura pós pandemia.

A segunda etapa consistiu em utilizar um modelo de rankeamento de fundos imobiliários e modelo de classificação.

Para fazer o rankeamento dos fundos, foi feita a estratégia de atribuir uma nota para cada um deles, que levará em consideração tanto os atributos atuais de cada um.

As variáveis e pesos utilizados foram:

- 'DividendYield': 0.02,
- 'DY (3M)Média': 0.08,
- 'DY (6M)Média': 0.15,
- 'DY (12M)Média': 0.25,
- 'P/VPA': 0.5

O porquê de utilizar tais variáveis foi pois tais variáveis são mais “conhecidas” do grande público, no entanto, as variáveis Dividend Yield, DY, e P/VPA, não devem, segundo especialistas no assunto de fundos imobiliários, ser utilizadas como únicos parâmetros para definição se um determinado fundo imobiliário é bom ou ruim, se vale ou não investir nesse fundo.

Essa recomendação dos especialistas em fundo imobiliário foi seguida, e para isso foram adotadas as seguintes medidas como hipótese da análise.

Essas foram as hipóteses:

- A utilização de 3 variáveis de DY, contendo as médias dos 3, 6 e 12 meses;

Essas outras hipóteses encontram-se como escopo negativo, pois não foram efetivamente finalizadas na presente versão do trabalho:

- Retirada da análise dos fundos monoativos;
- Retirada de fundos que tenham negociação (liquidez diária) menor que 1000, para garantir que o fundo não seja manipulado por alguns cotistas e para garantir que o fundo tenha representatividade na bolsa de valores;

- Retirada de valores NEGATIVOS de variáveis, tal qual rentabilidades negativas, que poderiam indicar que o fundo tem se depreciado ao longo do tempo.

Feita a atribuição das notas a todos os fundos imobiliários, foi necessário avaliar como está a distribuição das Notas e corrigir possíveis distorções.

E o método de correção foi um dos métodos amplamente utilizados para a normalização. O resultado da normalização subtrai o valor mínimo de um quadro de dados e divide-o pela diferença entre o valor mais alto e mais baixo da coluna correspondente.

Dado que todas as notas atribuídas aos FIs foram normalizadas, o modelo de classificação pode ser aplicado.

Levando em consideração a nota obtida de cada modelo, os FIs foram separados em 3 categorias em que queremos classificar: Ruim, Médio e Bom.

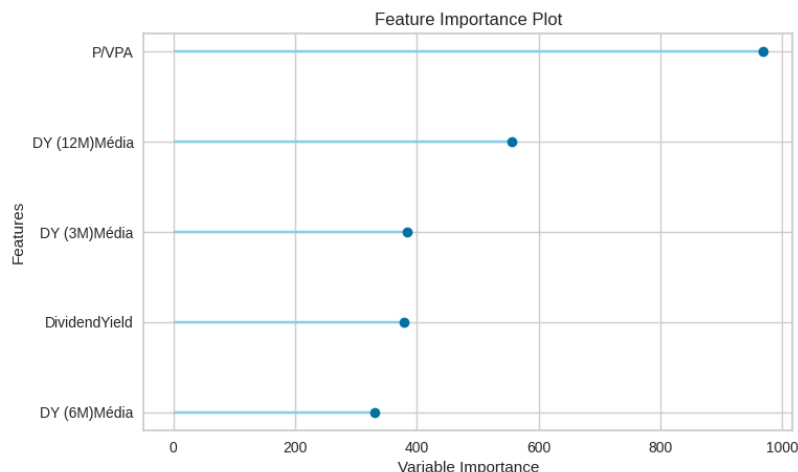
Foi então aplicada a biblioteca de Python, chamada Pycaret, para aplicar vários modelos de ML nos dados, descobrir qual o que melhor consegue classificá-los e para definir qual variável utilizada na análise, que é mais relevante para o modelo.

O uso da Pycaret, deu como resultado o que é mostrado na figura abaixo, e nesta figura encontra-se uma lista de algoritmos utilizados na pilha de treinamentos feitos pela biblioteca.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lr	Logistic Regression	0.8811	0.6769	0.7087	0.8519	0.8665	0.7327	0.7557	0.449
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
knn	K Neighbors Classifier	0.9118	0.6741	0.9369	0.9194	0.9159	0.8214	0.8310	0.134
lr	Logistic Regression	0.8811	0.6769	0.7087	0.8519	0.8665	0.7327	0.7557	0.449
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
knn	K Neighbors Classifier	0.9118	0.6741	0.9369	0.9194	0.9159	0.8214	0.8310	0.134
lr	Logistic Regression	0.8811	0.6769	0.7087	0.8519	0.8665	0.7327	0.7557	0.449
nb	Naive Bayes	0.8805	0.6578	0.9198	0.8755	0.8866	0.7430	0.7552	0.019
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
knn	K Neighbors Classifier	0.9118	0.6741	0.9369	0.9194	0.9159	0.8214	0.8310	0.134
dt	Decision Tree Classifier	0.8913	0.6091	0.9143	0.9020	0.8942	0.7755	0.7876	0.022
lr	Logistic Regression	0.8811	0.6769	0.7087	0.8519	0.8665	0.7327	0.7557	0.449
nb	Naive Bayes	0.8805	0.6578	0.9198	0.8755	0.8866	0.7430	0.7552	0.019
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
svm	SVM - Linear Kernel	0.9171	0.0000	0.7405	0.8828	0.9051	0.8168	0.8342	0.079
knn	K Neighbors Classifier	0.9118	0.6741	0.9369	0.9194	0.9159	0.8214	0.8310	0.134
dt	Decision Tree Classifier	0.8913	0.6091	0.9143	0.9020	0.8942	0.7755	0.7876	0.022
lr	Logistic Regression	0.8811	0.6769	0.7087	0.8519	0.8665	0.7327	0.7557	0.449
nb	Naive Bayes	0.8805	0.6578	0.9198	0.8755	0.8866	0.7430	0.7552	0.019
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
svm	SVM - Linear Kernel	0.9171	0.0000	0.7405	0.8828	0.9051	0.8168	0.8342	0.079
knn	K Neighbors Classifier	0.9118	0.6741	0.9369	0.9194	0.9159	0.8214	0.8310	0.134
dt	Decision Tree Classifier	0.8913	0.6091	0.9143	0.9020	0.8942	0.7755	0.7876	0.022
lr	Logistic Regression	0.8811	0.6769	0.7087	0.8519	0.8665	0.7327	0.7557	0.449
nb	Naive Bayes	0.8805	0.6578	0.9198	0.8755	0.8866	0.7430	0.7552	0.019
ridge	Ridge Classifier	0.8705	0.0000	0.6996	0.8437	0.8582	0.7141	0.7330	0.025
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
svm	SVM - Linear Kernel	0.9171	0.0000	0.7405	0.8828	0.9051	0.8168	0.8342	0.079
knn	K Neighbors Classifier	0.9118	0.6741	0.9369	0.9194	0.9159	0.8214	0.8310	0.134
rf	Random Forest Classifier	0.8966	0.6813	0.9246	0.8967	0.8978	0.7864	0.7979	0.641

A execução acima foi realizada diversas vezes, e nessa diversas vezes o algoritmo pelo rank da acurácia foi o GradientBoosting Classifier.

Já das variáveis utilizadas no treinamento a que mais influenciou no treinamento foi a P/PVA e o resultado foi o mostrado abaixo:



O resultado acima que mostrou em uma escala decrescente de importância P/VPA, DY(12)Média, DY(3)Média, DividendeYield e DY(6)Média.

Com os resultados acima, a execução do algoritmo GradientBoosting Classifier sobre a base, apresentou como resultado o mostrado na figura à baixo:

```

Fitting 10 folds for each of 50 candidates, totalling 500 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent work
[Parallel(n_jobs=-1)]: Done 62 tasks      | elapsed: 25.3s
[Parallel(n_jobs=-1)]: Done 212 tasks    | elapsed: 1.2min
[Parallel(n_jobs=-1)]: Done 462 tasks    | elapsed: 2.5min
[Parallel(n_jobs=-1)]: Done 500 out of 500 | elapsed: 2.7min finished

```

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Fold							
0	0.8000	0.9192	0.8690	0.8233	0.8034	0.6296	0.6418
1	0.9500	0.9899	0.9722	0.9562	0.9506	0.9052	0.9097
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0.8421	0.0000	0.8333	0.9091	0.8696	0.6705	0.6746
4	0.9474	0.0000	1.0000	0.9231	0.9600	0.8834	0.8895
5	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0.9474	1.0000	0.9444	0.9514	0.9460	0.8908	0.8967
7	0.8947	0.9422	0.9167	0.8947	0.8947	0.7889	0.7889
8	0.9474	0.9884	0.9444	0.9514	0.9460	0.8908	0.8967
9	0.9474	0.9769	0.9444	0.9514	0.9460	0.8908	0.8967
Mean	0.9276	0.6817	0.9425	0.9361	0.9316	0.8550	0.8595
Std	0.0610	0.4469	0.0537	0.0496	0.0573	0.1177	0.1156

Após o modelo ter sido definido e o treinamento ter sido realizados, a etapa de teste se deu sobre uma base de dados dos mesmos FIs, mas em um período de tempo diferente do treinamento.

```
1 print('Fundo IRDM11: ',best_model.predict(df_irdm.iloc[:,1:]))
2 print('Fundo VISC11: ',best_model.predict(df_bbpo.iloc[:,1:]))
3 print('Fundo ALZR11: ',best_model.predict(df_alzr.iloc[:,1:]))
4 print('Fundo CPTS11: ',best_model.predict(df_qiff.iloc[:,1:]))
5 print('Fundo XPLG11: ',best_model.predict(df_xpht.iloc[:,1:]))
6 print('Fundo VRTA11: ',best_model.predict(df_edga.iloc[:,1:]))
```

```
Fundo IRDM11: [0]
Fundo VISC11: [1]
Fundo ALZR11: [1]
Fundo CPTS11: [0]
Fundo XPLG11: [1]
Fundo VRTA11: [0]
```

Como a label 0 corresponde a bom, a label 1 corresponde a médio, e a label 2 corresponde a ruim, podemos ver que nosso modelo conseguiu classificar esses exemplos corretamente!

Ou seja, o modelo representou bem o cenário apresentado e pode ser aprimorado para melhor acurácia.

5. Registros de Homologação

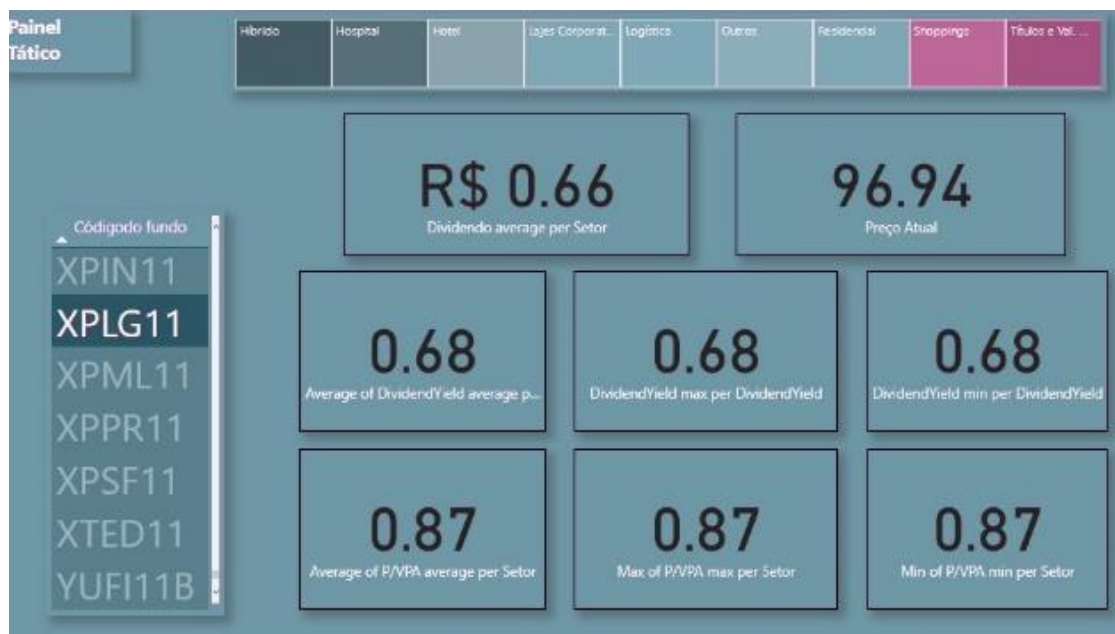
Para assegurar que os dados apresentados nos painéis estavam coerentes com a fonte de dados. Os dados mostrados nos painéis, publicados na interface web do PowerBI foram confrontados com os dados contidos na base fonte original, sem tratamento algum, existente na internet.

fundsexplorer							
CÓDIGO DO FUNDO	SETOR	PREÇO ATUAL	LIQUIDEZ DIÁRIA	DIVIDENDO	DIVIDEND YIELD	DY (3M) ACUMULADO	DY (6M) ACUMULADO
APCM11	Outros	R\$ 90,00	121100,0	R\$ 1,13	1,10%	3,40%	6,30%
XPCM11	Lajes Corporativas	R\$ 17,87	3831,0	R\$ 0,18	1,05%	2,75%	5,18%
XPTH11	Hotel	R\$ 98,00	80,0	R\$ 0,34	0,40%	1,05%	2,23%
XPIN11	Outros	R\$ 82,50	8001,0	R\$ 0,61	0,74%	2,14%	4,26%
XPLG11	Logística	R\$ 96,94	51748,0	R\$ 0,66	0,68%	2,04%	4,03%
XPML11	Shoppings	R\$ 98,38	125553,0	R\$ 0,64	0,71%	2,24%	3,93%
XPPT11	Outros	R\$ 59,12	21860,0	R\$ 0,45	0,80%	2,27%	4,87%
XPSF11	Outros	R\$ 69,63	13665,0	R\$ 0,67	0,96%	2,69%	5,41%
XTED11	Lajes Corporativas	R\$ 6,18	449,0	R\$ 0,00	0,00%	0,00%	0,00%

Sendo a linha selecionada em azul acima descrita completamente abaixo:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Código do fundo	Sector	Preço Atual	Liquidez Diária	Dividend Yield	DY (3M) Acumulado	DY (6M) Acumulado	DY (12M) Acumulado	DY (3M) Média	DY (6M) Média	DY (12M) Média	DY Ano	Variação Preço	Rentab. Período	Rentab. Acumulado	Patrimônio Líq.	VPA	PIVPA	DY Patrimonial	Variação Patrimonial	Rentab. Patrimonial Período	Rentab. Patrimonial Acumulado	Vacância Física	Vacância Financeira	Quantidade Ativos	
1	XPLG11	Logística	R\$ 96,68	48030,0	R\$ 0,66	0,68%	2,04%	4,03%	7,42%	0,68%	0,67%	0,62%	2,04%	5,82%	6,54%	0,97%	R\$ 3.021.396.048,41	R\$ 111,51	0,87	0,59%	1,49%	2,09%	2,69%	8,80%	1,20%	13

E os dados mostrados em um dos dashboard, no caso o Tático foram os seguintes:



- Considerações:

Pode haver em alguns casos em decorrência do processo de ETL no PowerBI, um lapso temporal, entre uma atualização no site fonte, uma vez que esse é atualizado diversas vezes ao dia e os dados mostrados nos dashboards, então neste caso se faz necessário atualizar os dos presentes nos dashboard.

No teste em questão a homologação foi feita após o fechamento do mercado em bolsa, ou seja, após as 18:00 horário de Brasília, para garantir a não alteração dos dados relativos aos Fundos Imobiliários durante os testes.

Se conclui que houve e há consistência nos dados mostrados nos painéis.

5. Conclusões

Primeiramente cabe ratificar que este trabalho em momento algum fez recomendação ou sugeriu a compra de algum ativo, pois não sou habilitado junto a Comissão de Valores Mobiliários (CVM) para tal.

Sobre o trabalho em si, este proporcionou a aplicação dos ensinamentos acumulados ao longo do curso de Pós-Graduação oferecido pela PUC-MG.

Ensinamentos especificamente na definição de indicadores, construção de *dashboards*, no aperfeiçoar o desenvolvimento na linguagem de programação Python e na utilização de modelos de *machine learning*, todos esses conhecimentos aplicados aos fundos imobiliários negociados na bolsa de valores brasileira.

Podemos enxergar que, como um todo, a situação dos FIIs, em 2019, estavam melhores que a situação atual e, com certeza, isso se deve pela situação econômica trazida pela crise sanitária do país com a COVID-19, no entanto, o cenário de 2022 mostra que há uma clara retomada aos patamares “normais” desse tipo de investimento.

Sendo essa retomada uma oportunidade de investimento, pois os valores e rentabilidade dos FIIs não atingiram seus patamares anteriores, o presente trabalho procurou balizar a escolha de que fundos imobiliários mereceriam uma atenção ou um estudo maior e que isso fosse mais uma variável nas análises próprias do investidor na decisão de investir ou não em um ativo.

6. Trabalhos Futuros

Como trabalhos futuros, a ideia é procurar meios que ajudem ainda mais a definir quais Fundo Imobiliários merecem mais atenção e deveriam ser estudados pelos interessados neste tipo de investimento. Não sendo isso uma recomendação de compra de ativo.

As melhorias que serão realizadas no desenvolvimento do presente projeto são:

1. Buscar a carteira de FIs propostas por casa de análises de investimento e confrontar as recomendações de fundos imobiliários e ter esses fundos como sendo os fundos analisados pode ser uma boa alternativa, para analisar ativos que já sofreram uma pré-análise qualitativa por analistas de investimentos credenciados.
2. Melhorar as definições de atribuição de pesos e aplicação de hipóteses de restrição de entrada de todos os FIs para restringir ainda mais a existência de outliers.
3. Avançar no estudo para aplicar um algoritmo de clusterização nos dados, e ver se eles são agrupados em conjunto, como CERTAMENTE haverá grupos, pois como já se tem os setores como sendo uma divisão macro dos FIs, o estudo seria para descobrir subgrupos. Pelo meu entendimento do mercado de FIs, baseado em alguns analistas deste mercado, um setor como o de Títulos, já existe subgrupos, e poderia ser dividido em High Yield, Middle Yield e High Grade , que pode ser entendida como uma classificação de risco X retorno do FI.

7. Links

Link 1. [Repositório do projeto]. Disponível em:

< https://github.com/lazs/tcc_pucmg >.

Link 2. [Fonte de dados]. Disponível em:

< <http://dados.cvm.gov.br/dataset/fi-cad> >.

Link 3. [Fonte de dados]. Disponível em:

< <https://www.kaggle.com/maverickjpa/fundos-imobiliarios-brasil-brazilian-funds-fiis>>.

REFERÊNCIAS

1. DUMONT, D. M.; RIBEIRO, J. A.; RODRIGUES, L. A. **Inteligência pública na era do conhecimento**. Rio de Janeiro: Revan, 2006. 329p.
2. FISCHMANN, A.; ZILBER, M. A. **Utilização de indicadores de desempenho como instrumento de suporte à gestão estratégica**, 1999.
3. FRANCISCHINI, P. G; FRANCISCHINI, A.S. N. **Indicadores de desempenho: dos objetivos à ação – métodos para elaborar KPIs e obter resultados**. Ed. Alta Books. Rio de Janeiro. 2017.
4. HEINZLE, R.; GAUTHIER, F. A. O.; FIALHO, F. A. P.; **Semântica nos sistemas de apoio à decisão: o estado da arte**. Revista da Unifebe, v. 1,p. Artigo 14, 2010.
5. KLUBECK, M. **Métricas: como melhorar os principais resultados de sua empresa**. Tradução: Eduardo Kraszczuk. São Paulo: Novatec Editora; New York, EUA: Apress Inc., 2012.
6. PARMENTER, D. **Key Performance Indicators: developing, implementing and using winning KPIs**. New Jersey: John Wiley & Sons, 2010.
7. POWER BI. **Plataforma de Análise e Business Intelligence**. Disponível em: <<http://powerbi.microsoft.com/>>. Acesso em: 11 Abril. 2022.
8. RIBEIRO, J. F. **Definição e implantação de KPIs para auxiliar a gestão de uma empresa de softwares**. Minas Gerais, 2017.
9. SIMON, H. A. **O processo decisório nas organizações**. Rio de Janeiro: Qualitymark, 2004. p.139-142.
10. TURBAN, E. [et al.]. **Business Intelligence: um enfoque gerencial para a inteligência do negócio**. Porto Alegre: Bookman, 2009. 253p.
11. ALBON, C. **Machine learning with Python cookbook: Practical solutions from preprocessing to deep learning**. O'Reilly Media, 2018.
12. AMARAL, F. **Introdução à ciência de dados: Mineração de dados e big data**. Alta Books, 2018.
13. BURGER, S. **Introduction to machine learning with R: Rigorous mathematical analysis**. O'Reilly Media, 2018.

14. DE CASTRO SILVA, D. **Introdução a mineração de dados**. Editora Saraiva, 2017. HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: Data mining, inference, and prediction**. Springer New York, 2013.
15. MARSLAND, S. **Machine learning: An algorithmic perspective**. CRC Press, 2011.
16. MÜLLER, A.; C, M.; GUIDO, S. **Introduction to machine learning with Python: A guide for data scientists**. O'Reilly Media, 2016.
17. OZDEMIR, S.; SUSARLA, D. **Feature engineering made easy: Identify unique features from your dataset in order to build powerful machine learning systems**. Packt Publishing, 2018.
18. RASCHKA, S. **Python machine learning**. Packt Publishing, 2015.