

Planejamento WCA



Inteligência Comercial

Manual de Procedimentos

Versão 0.02

	Responsável	Data
Elaborado	Alexsandro Augusto Ignácio	19/04/2021
Revisado		
Ajustado		
Aprovado		

Índice

Introdução.....	1
Objetivo	1
1 Levantamento de Requisitos.....	2
2 Construção do Blueprint	3
2.1 Cronograma do Projeto.....	3
3 Python	5
3.1 Bibliotecas	5
3.2 Documentação de Aplicações em Python – Read the Docs.....	8
4 Pyplan	9
4.1 Acesso ao Pyplan	9
4.2 Navegação	9
4.2.1 Diagrama de Influência	10
4.2.2 Interface e Menus	11
4.3 Modularização.....	12
5 Metodologias de Planejamento - Ciência de Dados	13
5.1 Análise Exploratória dos dados	13
5.2 Tipos de Variáveis.....	13
5.3 Modelos Estatísticos.....	14
5.4 Gráficos Estatísticos	21
Dicionário de Conteúdo	22
1.1 Palavras-Chave	22
A	22
B.....	22
C.....	22
D	22
E.....	23
F.....	23
G	23
H	23
I.....	23
J	23
K.....	23

L.....	23
M.....	23
N.....	23
O.....	24
P.....	24
Q.....	24
R.....	24
S.....	24
T.....	25
U.....	25
V.....	25
W.....	25
X.....	25
Y.....	25
Z.....	25
1.2 Fontes.....	25
A.....	25
B.....	25
C.....	25
D.....	25
E.....	25
F.....	25
G.....	25
H.....	25
I.....	25
J.....	25
K.....	25
L.....	25
M.....	25
N.....	25
O.....	26
P.....	26
Q.....	26
R.....	26
S.....	26
T.....	26

U	26
V	26
W	26
X.....	26
Y.....	26
Z.....	26
Conclusão	26

Introdução

Abordaremos neste documento os processos de estudos relativos à programação python e suas bibliotecas, a programação Pyplan® e suas aplicações em projetos de Data Analytics e Business Planning, utilização de modelos de séries temporais, estatísticos, machine learning e de performance que suportam todos os princípios da construção de modelos de S&OP, Demand Planning ou Forecast, Demand Seizing, Supply Chain Management, Sell in, Sell out, dentre outros de acordo com sua peculiaridade.

A área de planejamento deve suportar as melhores práticas a fim de proporcionar os melhores fluxos de trabalho, potencializar o negócio tanto na área comercial como também na área de operações, e promover a integração entre todos, sendo uma ponte nos processos de Inteligência Comercial 4.0:



Pyplan® | Planejamento Integrado de Demanda, S&OP e IBP.



SalesTech Intelligence® | Desdobramento de Metas de Vendas.



e-Decision Analytics® | Análise de Eficiência e Desempenho de Contexto.

Fonte logos: www.wca-ec.com.br

Objetivo

Este manual tem como premissa elucidar os métodos para a elaboração de soluções de negócio customizadas de acordo com a necessidade do cliente, onde os processos são mapeados e transcritos no Blueprint de acordo com os requisitos apresentados.

1 Levantamento de Requisitos

Para levantamento das funcionalidades da solução é necessário o levantamento de requisitos que são a representação dos processos realizados pelas empresas. Através deste podemos criar uma solução que elimine o ponto de dor que o cliente possui em sua estrutura de negócio. Este é feito em duas partes: entendimento do problema e definição da solução.

É fundamental o entendimento do core dos processos e eventos que afetam as necessidades dos usuários das áreas de S&OP e Planejamento.

Os Requisitos podem ser divididos em três categorias:

- Requisitos funcionais: abordam o que o sistema deve fazer.
- Requisitos não funcionais: são características de qualidade (performance, confiabilidade, portabilidade, segurança, usabilidade) que o sistema deve possuir, por exemplo o sistema deve atuar com determinada linguagem ou possuir certo grau de desempenho por requisição ou estar alocado em determinado servidor.
- Regras de negócio: são premissas e restrições aplicadas a uma operação comercial de uma empresa, que precisam ser atendidas para que o negócio funcione da maneira esperada. As regras de negócio definem como o sistema fará o atendimento às necessidades/exigências definidas; uma RN pode ser compreendida quanto a como um requisito funcional se realizará.

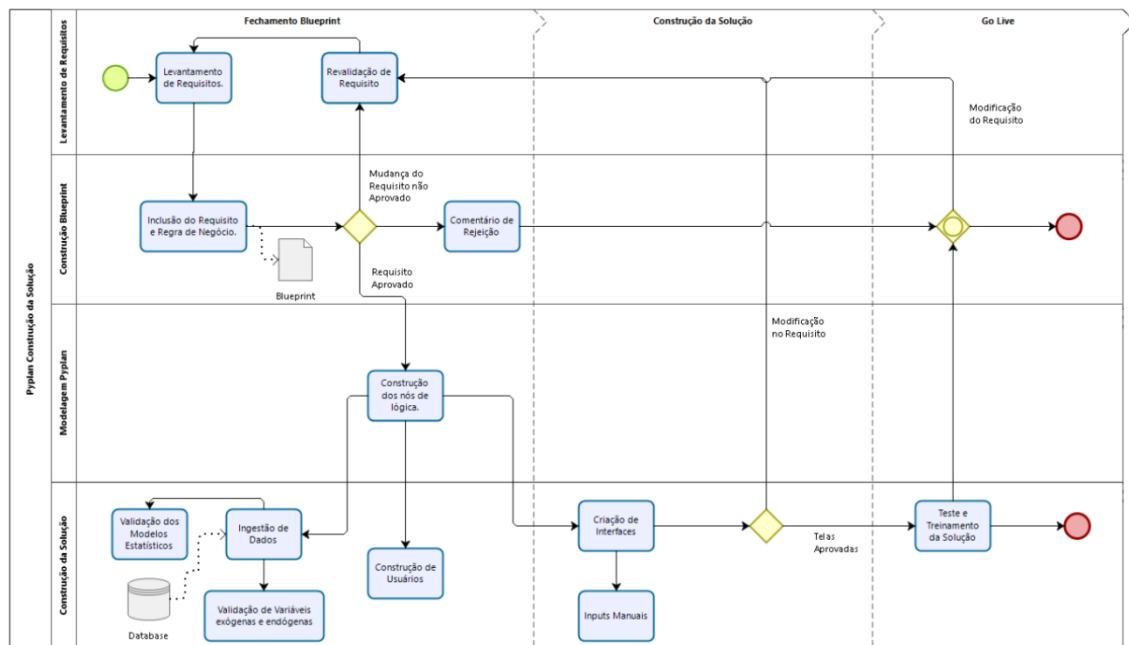


Fonte imagem: WCA

2 Construção do Blueprint

O Blueprint é a materialização do mapeamento dos requisitos e a descrição de como a solução deve se comportar e quais serão suas qualidades funcionais e não funcionais. Ele proporciona uma visão de quais funcionalidades, ações, tecnologias, telas, integrações, tempo de desenvolvimento e custo agregado são necessários para a entrega da solução ao cliente.

Visão atual do processo.



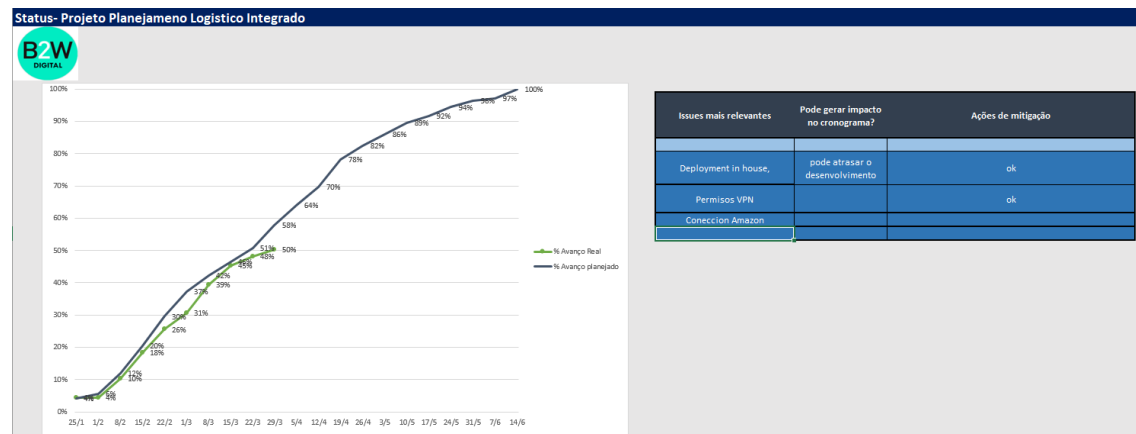
Fonte imagem: WCA

2.1 Cronograma do Projeto

A elaboração do Cronograma está atrelada as tarefas elaboradas no blueprint e ajudam a entender em que momento do projeto estamos e quais são os prazos de desenvolvimento necessários para cada tarefa e quais os ajustes para o go live dentro do tempo estipulado com o cliente para entrega da solução.

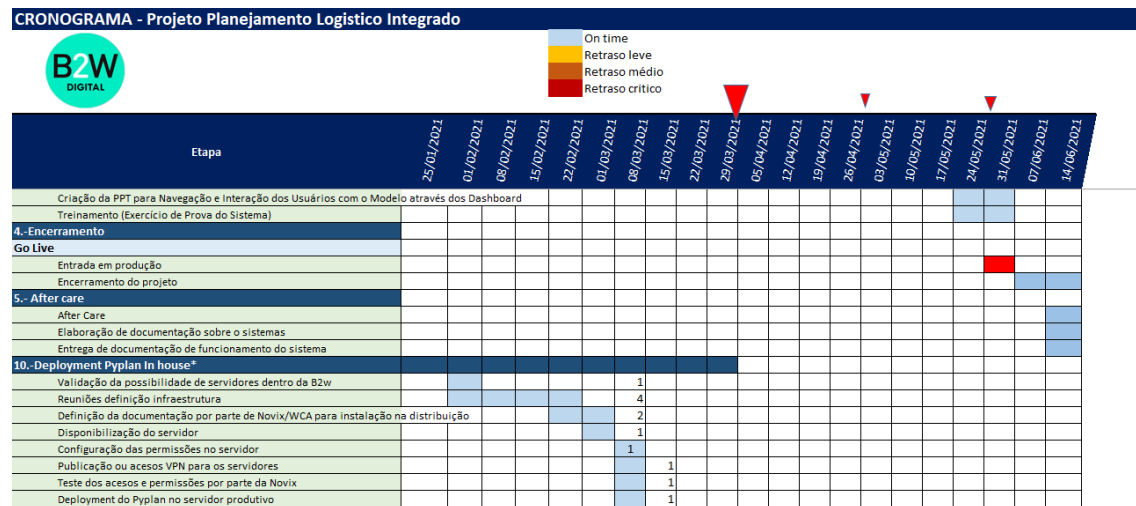
Exemplo do processo atual liderado pela Novix:

Projeção das Tarefas.



Fonte imagem: Novix

Tarefas Idealizadas no Blueprint.



Fonte imagem: Novix

3 Python

Python é uma linguagem amigável e de alto nível, com uma curva de aprendizagem muito maior, se comparada com outras tecnologias devido a sua simplicidade na escrita da sintaxe.

Possui uma comunidade ativa e com uma grande gama de bibliotecas que oferecem suporte para os mais diversos projetos, tais como:

- Desenvolvimento WEB.
- Data Science.
- Inteligência Artificial.
- Numérico e Científico.
- Educacional.
- Desenvolvimento de GUIs (Interfaces Gráficas).
- Desenvolvimento de Software e Automação de processos (testes e Scripts tasks).
- Aplicações de Negócios (Business Applications - ERP).

Dentre outras inúmeras aplicações que são desenvolvidas todos os dias dentro da comunidade python.



Fonte imagem: Python

3.1 Bibliotecas

PyPI é o gerenciador oficial de pacotes do python, é possível encontrar uma diversidade muito grande de soluções de algoritmos prontos para cada tipo de aplicação que existe hoje no mercado. Na área de planejamento utilizaremos os pacotes mais utilizado para Data Science, que são:

- Os e Sys - possibilita modificar e caminhar dentro do diretório do sistema operacional e utilizar comandos do CMD (prompt de comando).

Exemplo:

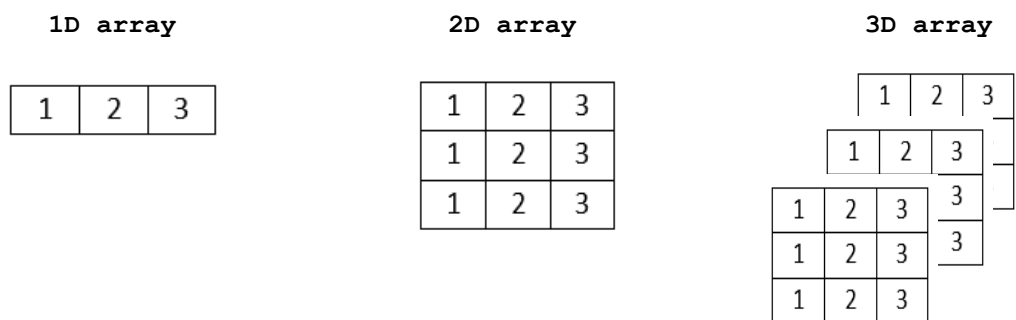
```
>>> import os  
>>> os.mkdir("C:\MyPythonProject")
```

- Numpy: Essa Biblioteca é a base para o ecossistema de data Science, é responsável pela vetorização e indexação de array (matrizes) n-dimensionais, oferece funções matemáticas abrangentes, geradores de números aleatórios, rotinas de álgebra linear, transformadas de Fourier e muito mais.

Exemplo:

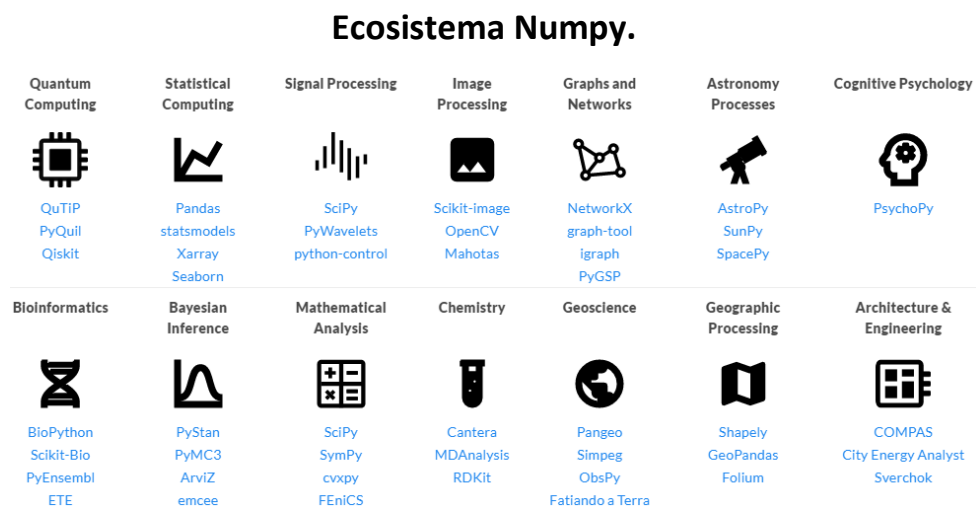
```
>>> import numpy as np
>>> x = np.arange(15, dtype=np.int64).reshape(3, 5)
>>> x[1:, ::2] = -99
>>> x
>>> array([[ 0,  1,  2,  3,  4],
          [-99,  6, -99,  8, -99],
          [-99, 11, -99, 13, -99]])
>>> x.max(axis=1)
array([ 4,  8, 13])
```

Exemplo das dimensões que um data array pode possuir:



Fonte imagem: WCA

O numpy distribui funcionalidades para uma grande gama de outras bibliotecas que se apropriam da sua capacidade de lidar com arrays multidimensionais.



Fonte imagem: Numpy

- Pandas:** Utiliza a biblioteca Numpy para manipulação de dados no formato de data frame, muito parecido com uma planilha do Excel. Possui grande utilização em machine learning e permite diversas operações de álgebra relacional, como projeção, junção, e concatenação, funções de limpeza, como por exemplo o preenchimento, substituição ou inserção de valores nulos (null) e comandos sql como junção de tabelas. Tem como destaque a criação de um objeto DataFrame rápido e eficiente para manipulação de dados com indexação integrada, pode ler e gravar dados em diversos formatos (csv, txt, xlsx, hdf5), possui alinhamento automático baseado em rótulo e em cálculos, fácil manipulação dos dados confusos de forma ordenada, possibilita a remodelagem e rotação flexível dos conjuntos de dados, as colunas podem ser inseridas e excluídas de estruturas de dados para mutabilidade de tamanho, é possível utilizar a funcionalidade da série temporal para a geração de intervalo de datas e conversão de frequência, estatísticas de janela móvel, mudança de data e atraso. Crie até mesmo compensações de tempo específicas de domínio e junte séries temporais sem perder dados.

Data Frame

		Colunas	Colunas	Colunas	Colunas
		NOME	PRODUTO	PREÇO	DATA
Linhas	0	Mesa	1654	1750.00	01/01/2021
Linhas	1	Cadeira	265	245.00	15/02/2021
Linhas	2	Sofá	164	3500.00	30/01/2021
Linhas	3	Cômoda	98	645.00	14/03/2021
Linhas	4	Banco	56	120.00	30/03/2021
Linhas	5	Armário	2235	2564.00	01/04/2021
		Series	Series	Series	Series

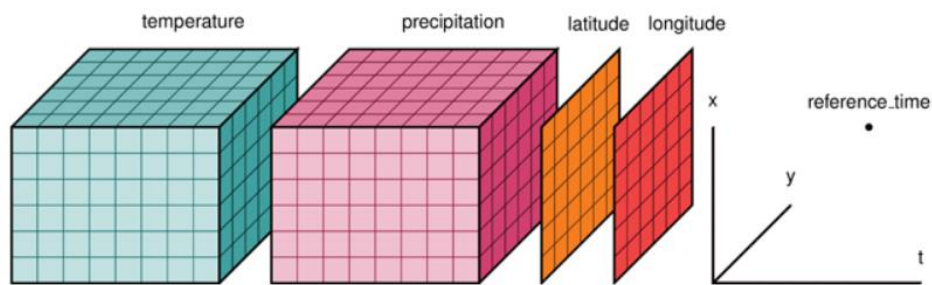
```
df = pd.DataFrame(data = {'Fulano': [8, 10, 4, 8, 6, 10, 8],
                          'Beltrano': [10, 2, 0.5, 1, 3, 9.5, 10],
                          'Sicrano': [7.5, 8, 7, 8, 8, 8.5, 7]},
                  index = ['Matemática',
                          'Português',
                          'Inglês',
                          'Geografia',
                          'História',
                          'Física',
                          'Química'])
df.rename_axis('Matérias', axis = 'columns', inplace = True)
df
```

Matérias	Fulano	Beltrano	Sicrano
Matemática	8	10.0	7.5
Português	10	2.0	8.0
Inglês	4	0.5	7.0
Geografia	8	1.0	8.0
História	6	3.0	8.0
Física	10	9.5	8.5
Química	8	10.0	7.0

Fonte imagem: WCA

- Xarray:** Apresenta rótulos na forma de dimensões, coordenadas e atributos no topo de matrizes multidimensionais semelhantes a NumPy, o que permite uma experiência de desenvolvedor mais intuitiva, concisa e menos sujeita a erros.

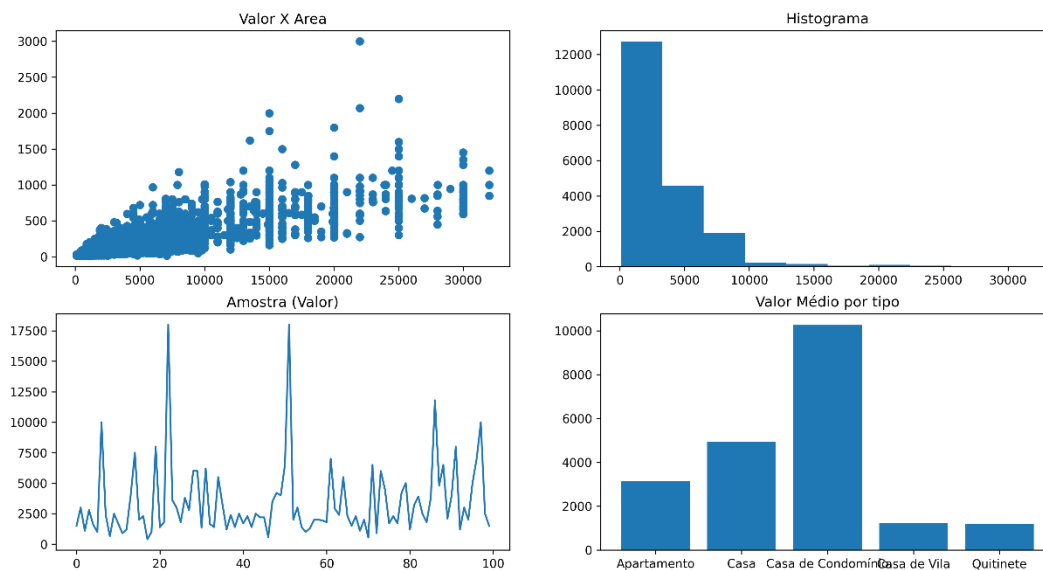
Xarray



Fonte imagem: Open research software

- Matplotlib e Seaborn: São duas bibliotecas que oferecem suporte a gráficos para a visualização de dados na linguagem Python. Fornece uma interface de alto nível para desenhar gráficos estatísticos atraentes e informativos.

Vizualização de Dados



Fonte imagem: WCA

3.1.1 Pacotes ou Packages

3.2 Documentação de Aplicações em Python – Read the Docs

4 Pyplan

O Pyplan é um ambiente de desenvolvimento Python destinado a criar e implementar facilmente aplicativos de análise de dados. Diferente de um Jupyter Notebook, onde o código é estruturado como uma lista de frases, no código Pyplan está contido em nós que funcionam como etapas de cálculo, organizados em um diagrama de influência hierárquica. Os nós podem ser avaliados e seu resultado inspecionado por meio de uma saída de console ou renderizado como uma tabela ou gráfico. As interfaces de usuário são criadas arrastando os nós em uma tela.

Entre seus recursos mais importantes, você encontrará:

- Programação gráfica assistida de arrastar e soltar.
- Diagrama de influência visual para representar o fluxo lógico.
- Criação fácil de interfaces de usuário interativas.
- Colaboração capacitada por meio da publicação e compartilhamento de um clique.
- Seguro e escalonável com padrões corporativos.
- Implantação na nuvem ou local.

4.1 Acesso ao Pyplan

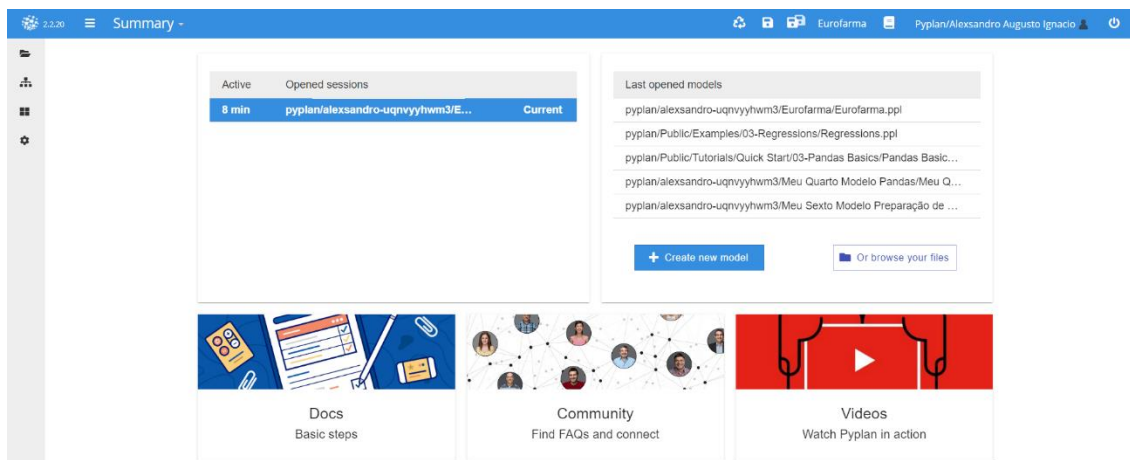
O acesso pode ser realizado via o próprio site pyplan.org que fornece uma versão gratuita para degustação e aprendizagem com login pelo linkedin. Outras formas de se realizar o acesso é via instalação no próprio sistema operacional ou via cadastramento de usuário no servidor de deployment no modelo SAAS.

4.2 Navegação

Sua IDE permite uma fácil navegação entre recursos que o sistema dispõe, desde a criação dos diagramas de influência, gerenciar os arquivos pelo file manager, criar e gerenciar as interfaces que realizam a visualização do processamento dos dados e lógicas aplicadas nos nós do diagrama de influência de cada projeto. Possui uma Biblioteca com os principais modelos de negócios e estáticos.

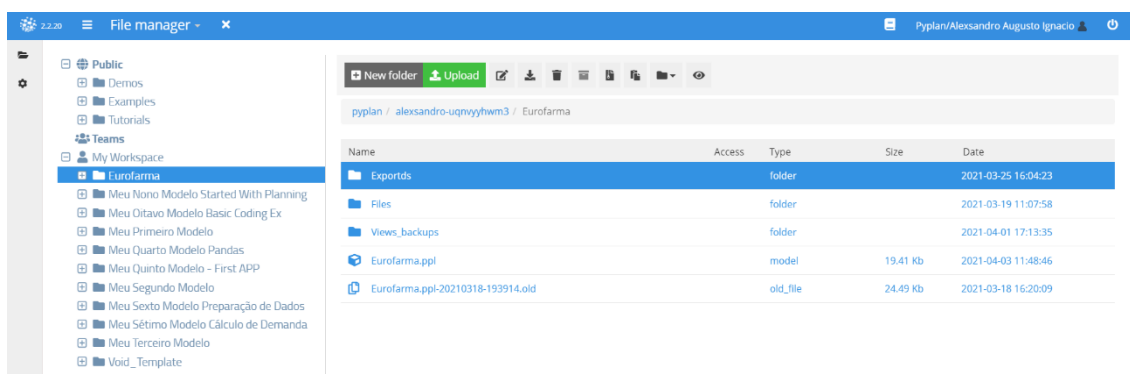
Do lado direito quando o modelo está aberto podemos acessar a barra de edição do diagrama, configuração do modelo e configuração de bibliotecas do python.

Painel de Navegação



Fonte imagem: WCA/Novix/Pyplan

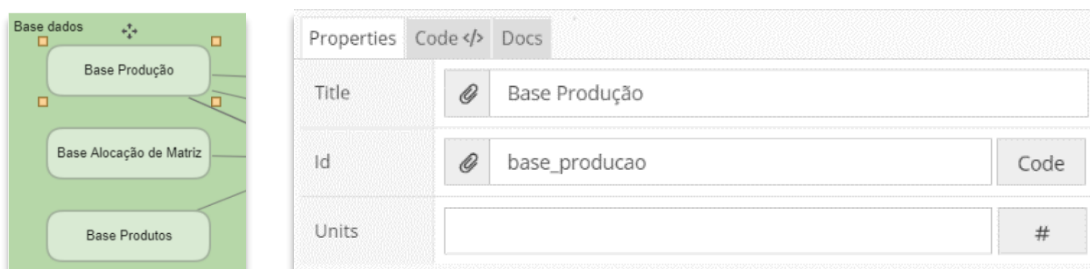
Gerenciador de Arquivos – File Manager



Fonte imagem: WCA/Novix/Pyplan

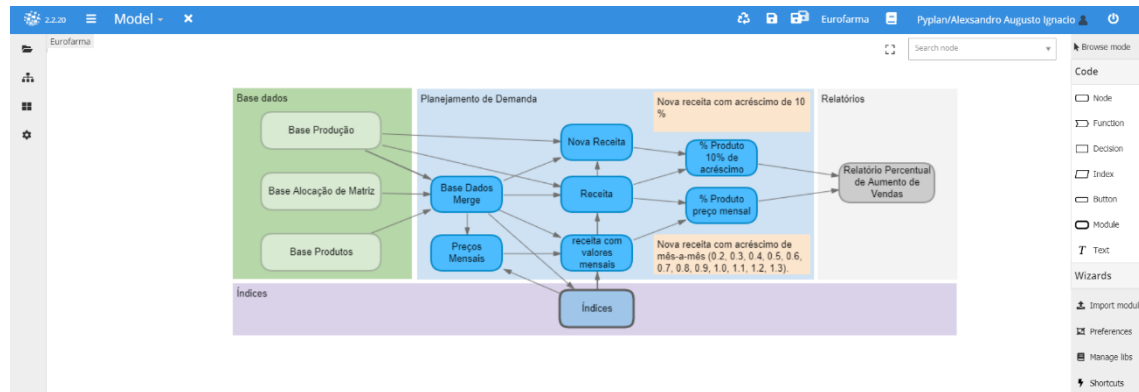
4.2.1 Diagrama de Influência

O Diagrama de influência no pyplan funciona com nós que recebem sempre uma lógica representada pela codificação em python. Esses nós são interligados pela entrada e saída de dados representados por id que são identificadores únicos daquele ponto específico:



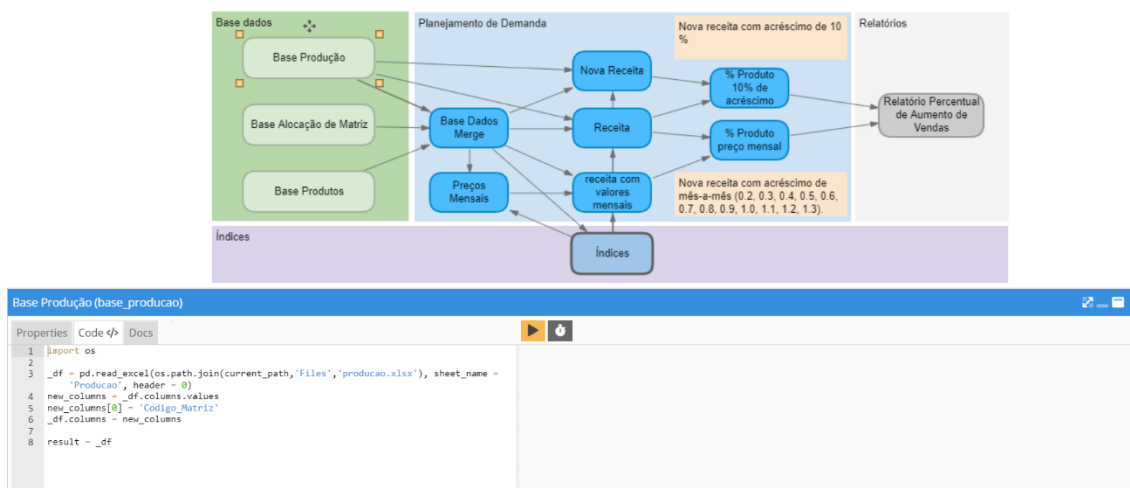
Fonte imagem: WCA/Novix/Pyplan

Modelo de Diagrama



Fonte imagem: WCA/Novix/Pyplan

Painel de Codificação



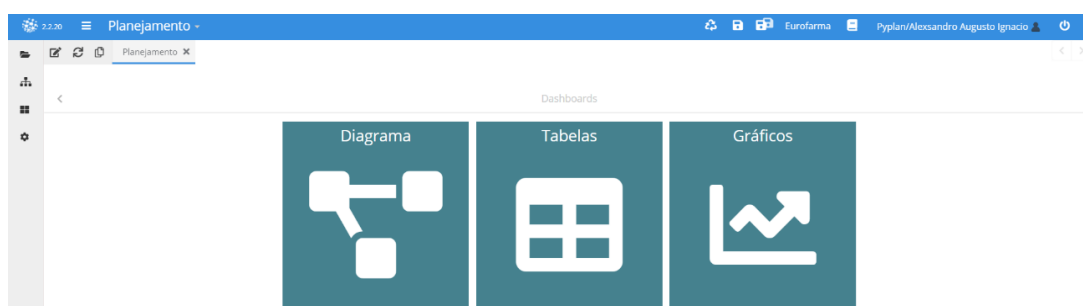
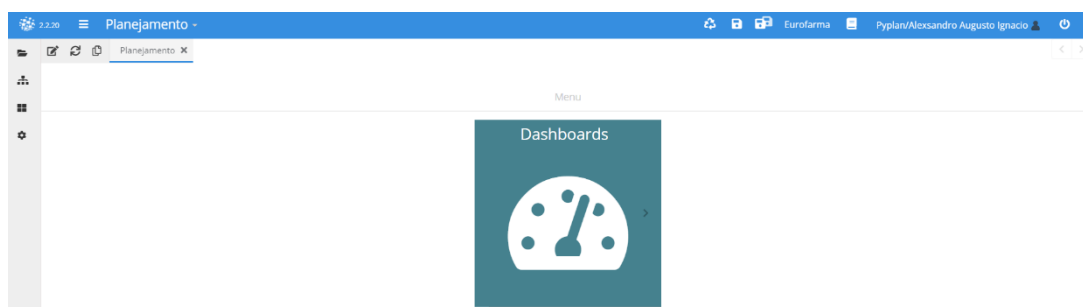
Fonte imagem: WCA/Novix/Pyplan

4.2.2 Interface e Menus

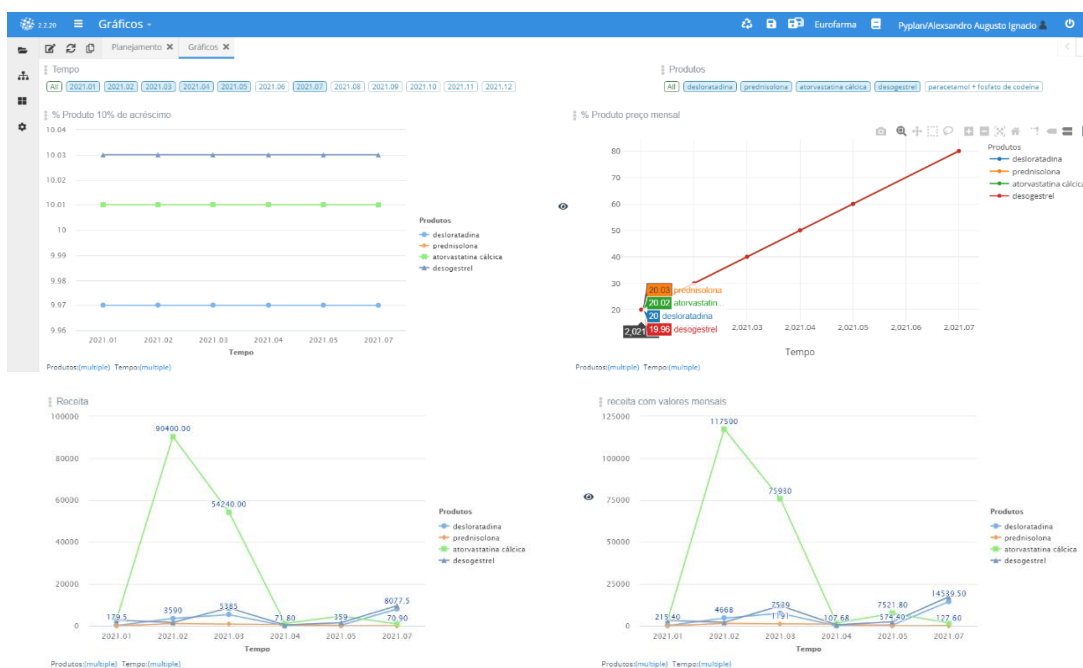
O pyplan oferece um recurso de interfaces que possibilita ao usuário interagir com os dados através de inputs manuais e seletores, podendo assim criar um dinamismo na forma de exibição dos dados em forma de gráficos ou tabelas.

Possibilita a organização destas interfaces em menus e submenus aumentando a sensação de versatilidade na usabilidade do sistema.

Menu



Interface



Fonte imagem: WCA/Novix/Pyplan

4.3 Modularização

- Lógica de Cálculos
- Regra de Negócios

5 Metodologias de Planejamento - Ciência de Dados

Neste documento temos como intenção demonstrar as principais metodologias aplicadas na ciência de dados e como elas se correlacionam com o pyplan e os métodos utilizados dentro de nossos clientes como forma de planejar e mensurar suas estratégias de negócio para o passado, presente e futuro.

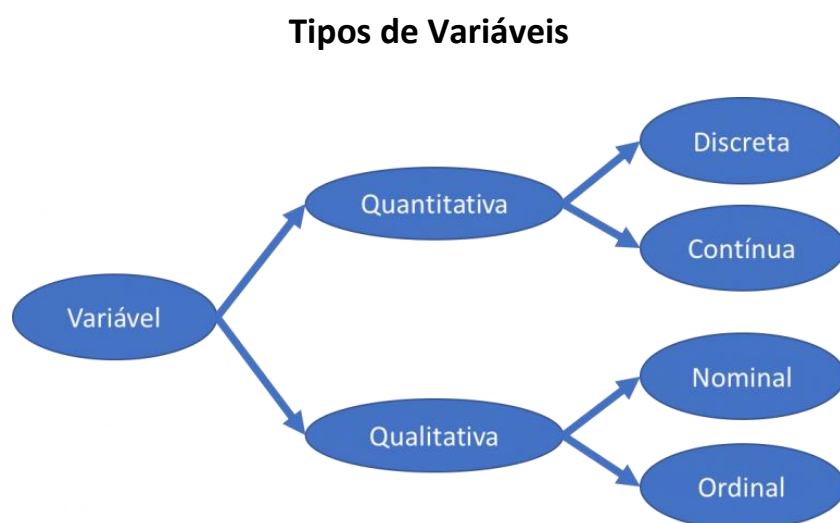
5.1 Análise Exploratória dos dados

A análise exploratória de dados consiste na junção de ciências matemáticas, estatísticas e programação, no processo de modelagem buscamos conhecer nossa base de dados e como eles estão relacionados para criar informações que agregam valor.

Esses modelos têm como premissa proporcionar uma visão preditiva do futuro e de ações sugestivas com base na informação gerada.

5.2 Tipos de Variáveis

As variáveis são valores que representam determinadas características dentro de uma pesquisa. Esses valores variam de elemento para elemento. É a medida em cada elemento da amostra ou população. Essas variáveis podem ter valores numéricos ou não numéricos e são classificadas assim:



Fonte imagem: WCA

Variáveis Quantitativas: são representadas por meio de números resultantes de uma contagem ou mensuração. Elas podem ser de dois tipos:

- Variáveis discretas: os valores representam um conjunto finito ou enumerável de números, e que resultam de uma contagem, por exemplo: Número de filhos (0,1,2,...), número de bactérias por amostra, número de copos de cerveja tomados por dia.
- Variáveis contínuas: os valores pertencem a um intervalo de números reais e representam uma mensuração como por exemplo altura ou peso de uma pessoa. Nesses casos números fracionais fazem sentido. Exemplo: tempo (relógio) e pressão arterial.

Variáveis Qualitativas: representam uma qualidade (ou atributo) de um indivíduo pesquisado, são definidas por várias categorias. São características que não possuem valores quantitativos. Essas variáveis podem ser de dois tipos:

- Variável nominal: quando não existe nenhuma ordenação nas possíveis representações. Exemplos: sexo, cor dos olhos, cor do cabelo, fumante/não fumante.
- Variável ordinal: quando apresentam uma ordem nos seus resultados. Exemplos: escolaridade (1, 2, 3 graus), mês de observação (janeiro, fevereiro, ..., dezembro.)

Variáveis exógenas e endógenas:

- Variável exógena refere-se a uma variável que é determinada fora do modelo e representa as entradas de um modelo.
- Variáveis endógenas são determinadas dentro do modelo e, portanto, representam as saídas de um modelo. O modelo especificado com as variáveis mostra como a mudança de uma variável exógena afeta todas as variáveis endógenas.

5.3 Modelos Estatísticos

Modelos estatísticos são a construção de hipóteses a partir da análise de dados, de sua relação e de outras variáveis para prever ou comprovar fatores futuros.

Distribuição de Frequências

- **Distribuição:** é o entendimento de como os dados se distribuem e se comportam dentro de uma variável de forma percentual.

Catégorias de Renda	Frequência
A	608
B	2178
C	7599
D	16700
E	49755

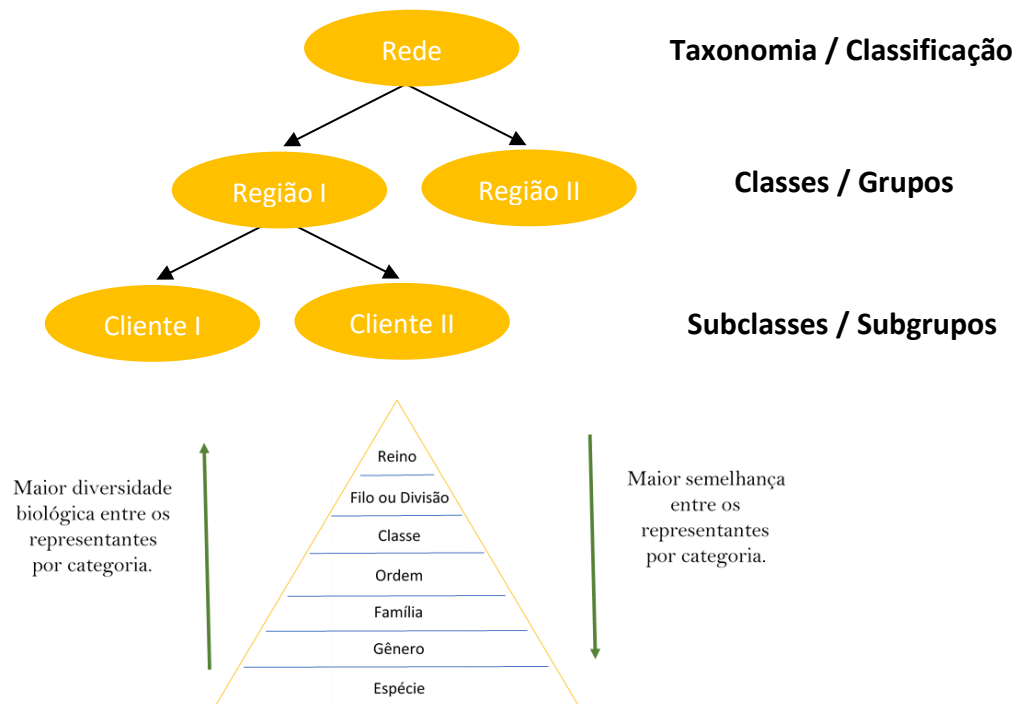
Fonte imagem: WCA

- **Frequência:** é visualização da repetição dentro daquela variável e o grau de sua participação dentro daquela amostra ou população.

Catégorias de Renda	Porcentagem (%)
A	0.01
B	0.03
C	0.10
D	0.22
E	0.65

Fonte imagem: WCA

- **Classificação:** ação ou efeito de classificar, de reunir em classes e em grupos respectivos, segundo um sistema ou método. [Biologia] Distribuição sistemática em diversas categorias segundo as analogias e caracteres comuns: classificação dos seres vivos.



Fonte imagem: WCA

- **Categorização:** é o processo pelo qual ideias e objetos são reconhecidos, diferenciados e classificados, e consiste em organizar os objetos de um dado universo em grupos ou categorias, com um propósito específico.



Fonte imagem: WCA

- **Regra de Sturges:** é um critério usado para determinar o número de classes ou intervalos necessários para representar graficamente um conjunto de dados estatísticos.

$$k = 1 + \frac{10}{3} \log_{10} n$$

```
>>> n = dados.shape[0] = 76840
>>> int(round(k = 1 + (10/3) * np.log10(n)))
>>> 17 faixas de distribuição.
```

Faixas (Regra de Sturges)	Frequência	Porcentagem (%)
-200.001 a 11764.706	75594	0.98
11764.706 a 23529.412	1022	0.01
23529.412 a 35294.11	169	0.00
35294.118 a 47058.824	19	0.00
47058.824 a 58823.529	16	0.00
58823.529 a 70588.235	5	0.00
70588.235 a 82352.941	4	0.00
82352.941 a 94117.647	1	0.00
94117.647 a 105882.353	6	0.00
105882.353 a 117647.059	0	0.00
117647.059 a 129411.765	1	0.00
129411.765 a 141176.471	0	0.00
141176.471 a 152941.176	0	0.00
152941.176 a 164705.882	0	0.00
164705.882 a 176470.588	0	0.00
176470.588 a 188235.294	0	0.00
188235.294 a 200000.000	3	0.00

Fonte imagem: WCA

Média e Mediana

- **Média:** soma dos valores dividido pela sua quantidade.

$$Média = \frac{\sum x}{n_x}$$

Matérias	Fulano	Beltrano	Sicrano
Matemática	8	10.0	7.5
Português	10	2.0	8.0
Inglês	4	0.5	7.0
Geografia	8	1.0	8.0
História	6	3.0	8.0
Física	10	9.5	8.5
Química	8	10.0	7.0

Fonte imagem: WCA

Média com python:

```
>>> media = (8 + 10 + 4 + 8 + 6 + 10 + 8) / 7
>>> round(media, 2)
>>> 7.71
```

Média com pandas:

```
>>> df['Fulano'].mean().round(2)
>>> 7.71
```

- **Mediana:** é um conjunto de informações numéricas onde o valor central corresponde à mediana desse conjunto. Dessa forma, é importante que esses valores sejam colocados em ordem, seja crescente ou decrescente. Se houver uma quantidade ímpar de valores numéricos, a mediana será o valor central do conjunto numérico. Se a quantidade de valores for um número par, devemos fazer uma média aritmética dos dois números centrais, e esse resultado será o valor da mediana.

1 - Organizar os elementos do conjunto de dados do menor para o maior.

Conjunto par = [1, 6, 5, 4] --> [1, 4, 5, 6]

Conjunto ímpar = [10, 30, 1, 548, 12] --> [1, 10, 12, 30, 548]

2 – No conjunto de dados é necessário identificar qual é o elemento (n) médio e se ele é par ou ímpar para definir a fórmula adequada.

$$\text{Ímpar: } Elemento = \frac{n+1}{2}$$

Elemento = 12

$$\text{Par: } Elemento = \frac{n}{2}$$

Elemento = 4

3 – Identifica o elemento mediano.

$$\text{Ímpar: } Mediana = X_{Elemento}$$

Mediana = 12

$$\text{Par: } Mediana = \frac{X_{Elemento1} + X_{Elemento2}}{2}$$

Mediana = 4.5

- **Moda:** é o valor mais frequente de um conjunto de dados. Dentro de dataset podemos possuir mais de uma moda, então chamamos de bimodal ou multimodal.

$$Mo = Li + \frac{D_1}{D_1 + D_2} \times h$$

Matérias	Fulano	Beltrano	Sicrano
Matemática	8	10.0	7.5
Português	10	2.0	8.0
Inglês	4	0.5	7.0
Geografia	8	1.0	8.0
História	6	3.0	8.0
Física	10	9.5	8.5
Química	8	10.0	7.0

Moda			
Matérias	Fulano	Beltrano	Sicrano
	8	10.0	8.0

Pynton/pandas:

```
>>> df.mode()
>>> Fulano = 8, Beltrano = 10 e Sicrano = 8
```

Fonte imagem: WCA/Alura

- **Medidas Separatrizes:** são números que dividem a sequência ordenada de dados em partes que contêm a mesma quantidade de elementos da série.

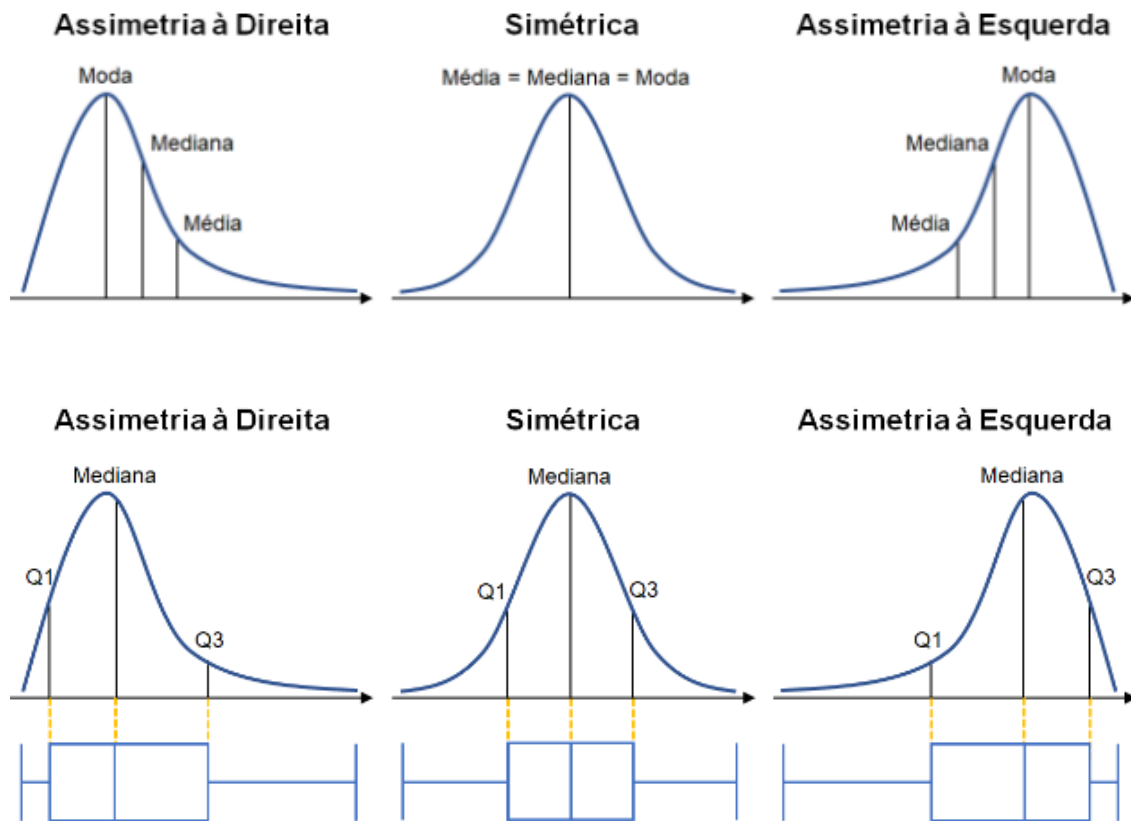
$$Pi = \frac{i(n + 1)}{100}$$

- **Mediana:** dividi a distribuição em duas partes iguais.
- **Quartis:** dividir a distribuição em quatro partes iguais.
- **Decis:** dividir a distribuição em dez partes iguais.
- **Percentis:** dividir a distribuição em cem partes iguais.

Relação entre média, mediana e moda

Podemos verificar como fica visível as tendências dos dados de acordo com as metodologias estatísticas que são aplicadas ao modelo.

Tendências



Fonte imagem: WCA/Alura

Medidas de Dispersão

Embora as medidas de posição forneçam uma sumarização bastante importante dos dados, elas podem não ser suficientes para caracterizar conjuntos distintos, especialmente quando as observações de determinada distribuição apresentarem dados muito dispersos.

- **Desvio médio absoluto:** conjunto de dados é a média das distâncias entre cada dado e a média.

$$DM = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}|$$

- **Variância:** é construída a partir das diferenças entre cada observação e a média dos dados, ou seja, o desvio em torno da média. No cálculo da variância, os desvios em torno da média são elevados ao quadrado.

Variância populacional:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

Variância amostral:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Matérias	Fulano	Desvio	Desvio	(Desvio)^2
Matemática	8	0.285714	0.285714	0.081633
Português	10	2.285.714	2.285.714	5.224.490
Inglês	4	-3.714.286	3.714.286	13.795.918
Geografia	8	0.285714	0.285714	0.081633
História	6	-1.714.286	1.714.286	2.938.776
Física	10	2.285.714	2.285.714	5.224.490
Química	8	0.285714	0.285714	0.081633

Variância = 4.57142857142857

Fonte imagem: WCA

- **Desvio Padrão:** é uma medida que expressa o grau de dispersão de um conjunto de dados. Ou seja, o desvio padrão indica o quanto um conjunto de dados é uniforme. Quanto mais próximo de 0 for o desvio padrão, mais homogêneo são os dados.

Desvio padrão populacional:

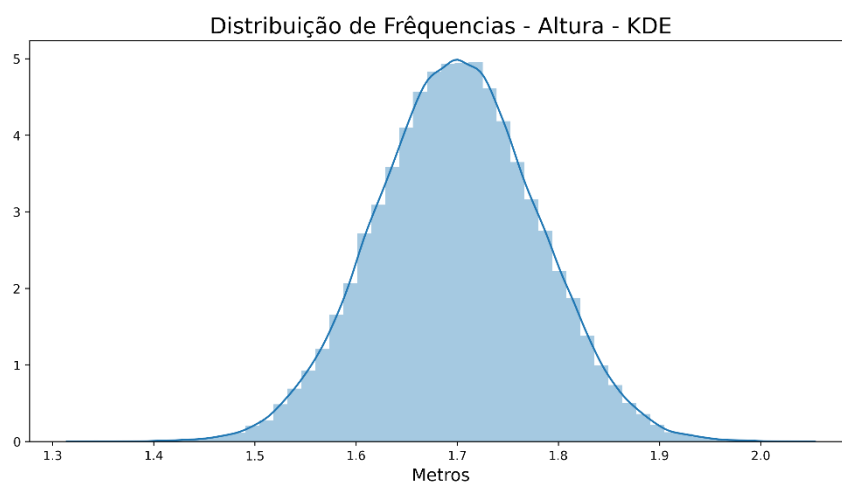
$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2} \Rightarrow \sigma = \sqrt{\sigma^2}$$

Desvio padrão amostral:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \Rightarrow S = \sqrt{S^2}$$

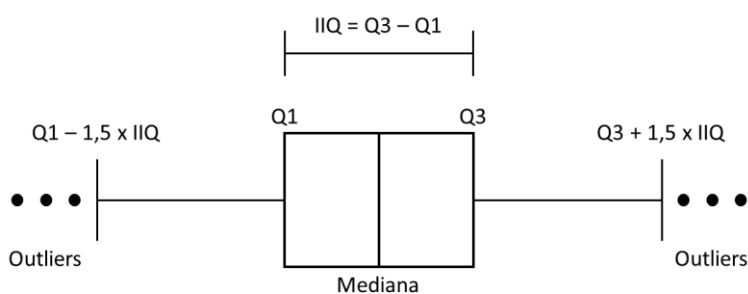
5.4 Gráficos Estatísticos

Histograma: é um gráfico de barras que demonstra uma distribuição de frequências, onde a base de cada uma das barras representa uma classe, e a altura a quantidade ou frequência absoluta com que o valor da classe ocorre.



Fonte imagem: WCA

Boxplot: proporciona uma ideia da posição, dispersão, assimetria, caudas e dados discrepantes (outliers). A posição central é dada pela mediana e a dispersão por IIQ. As posições relativas de Q1, Mediana e Q3 dão uma noção da simetria da distribuição. Os comprimentos das caudas são dados pelas linhas que vão do retângulo aos valores remotos e pelos valores atípicos.



Box-plot

Fonte imagem: WCA

Dicionário de Conteúdo

Com o objetivo de empoderar os leitores deste manual, segue lista de palavras chaves e fontes de dados do qual este documento foi constituído.

1.1 Palavras-Chave

A

- **Análise:** método de pensamento voltado para a compreensão ou explicação de um fenômeno complexo, que consiste em reduzir uma realidade intrincada, de difícil apreensão global, em seus componentes básicos e simples.
- **Apache Spark:** é um mecanismo de análise unificado para processamento de dados em grande escala. Ele fornece APIs de alto nível em Java, Scala, Python e R e um mecanismo otimizado que oferece suporte a gráficos de execução geral.

B

- **B2B:** é a sigla para o termo em inglês “Business to Business”, e a expressão se refere às empresas que vendem para outras empresas.
- **B2C:** é a sigla para Business to Consumer, ou empresa para consumidor, que representa as transações entre empresas e consumidores finais.
- **Budget:** traça a estimativa de vendas, custos e despesas de uma companhia para um determinado período de tempo.
- **Business Planning:** é um plano de negócio que traça um caminho, uma direção para alcançar os resultados quantitativos e qualificativos esperado pelo board. Este plano especifica quais metodologias serão usadas, em quais prazo serão alcançados e quais metas precisam ser conquistadas no decorrer desse caminho.

C

- **CDs:** Centro de Distribuições são espaços para armazenar mercadorias destinado ao recebimento, à separação e ao envio de produtos.
- **Canal direto:** É uma unidade de trabalho que pertence à mesma empresa de produção e que é diretamente responsável pela entrega das mercadorias aos seus clientes. Isso requer que a empresa tenha espaço para armazenar os produtos e os meios de transporte para realizar as entregas. Além disso, deve ter contato direto com os clientes.
- **Canal indireto:** São empresas terceiras responsáveis por fornecer o serviço. Tais distribuidoras dividem o trabalho de acordo com custos, preços, tipo de mercado e forma de comercialização do produto.
- **Canal Híbrido:** é uma mistura entre os canais direto e indireto.

D

- **Desdobramento de Metas:** é a tradução dos objetivos macro da empresa em valores ou quantidades segmentados de forma que traduzem para os departamentos / unidades de negócio quais são seus objetivos (micro) individuais de atingimento. Há várias metodologias para acompanhamento, exemplo: Balanced Scorecard - BSC, Objective and Key Results - OKRs e acompanhamento dos resultados.

- **Deployment:** Implantação.
- **Demand Seizing (Detecção de Demanda):** capacidade de projetar a demanda de acordo com sua movimentação de mercado, movimentação de perfis de clientes intermediário (B2B) e finais (B2C).
- **Demand Planning (Planejamento de Demanda):** capacidade de planejar, projetar e coordenar os fluxos de informações de produtos, insumos e serviços que devem estar a disposição para comercialização.
- **Data Analytics:** é o processo de analisar informações (dados) com um propósito específico. Isto é, pesquisar e responder perguntas com base em dados e com uma metodologia clara para todos os participantes.
- **Datasets:** um conjunto de dados ou "dataset" é uma coleção de dados normalmente tabulados. Por cada elemento se indicam várias características. Cada coluna representa uma variável particular. Cada linha corresponde a um determinado membro do conjunto de dados em questão. Cada valor é conhecido como um dado.
- **DataFrame:** é semelhante a uma matriz mas as suas colunas têm nomes e podem conter dados de tipo diferente. Pode ser visto como uma tabela de uma base de dados, em que cada linha corresponde a um registo (linha) da tabela. Cada coluna corresponde às propriedades (campos) a serem armazenadas para cada registo da tabela.
- **Desempenho:** cumprimento de obrigação ou de promessa; execução.
- **Days of Inventory on Hand:** é a cobertura dos estoques (Quanto tempo demora para uma empresa transformar seus estoques em vendas).
- **Drill-down:** detalhamento dos dados dentro de uma hierarquia.

E

- **Eficiência:** virtude ou característica de (alguém ou algo) ser competente, produtivo, de conseguir o melhor rendimento com o mínimo de erros e/ou dispêndios.

F

- **Fair Share:** consiste em distribuir o estoque de forma que os CDs / Brokers tenham um nível de cobertura semelhante.
- **Forecast (Previsão ou Planejamento):** O forecast é o orçamento ajustado de uma empresa. O objetivo da elaboração de um forecast empresarial é garantir que sejam cumpridas as metas estipuladas no orçamento inicial.

G

H

I

J

K

L

- **Lead times:** dias de demora entre cargas geradas e despachadas e dias de demora entre quando elas deixam as Fábricas ou CDs até que cheguem no seu destino.

M

N

- **Net Sales:** Vendas líquidas.

O

- **Outliers:** são dados que se diferenciam drasticamente de todos os outros, são pontos fora da curva normal (o que é curva normal?). Em outras palavras, um outlier é um valor que foge da normalidade e que pode (e provavelmente irá) causar anomalias nos resultados obtidos por meio de algoritmos e sistemas de análise.

P

- **POC:** Proof of concept ou Prova de Conceito, é um termo utilizado para denominar um modelo prático que possa provar o conceito estabelecido.
- **Protótipo:** é um modelo preliminar, são usados para aumentar a chance de sucesso do projeto.
- **Product Sales:** faturamento líquido de impostos.
- **Parquet:** é um formato colunar compatível com muitos outros sistemas de processamento de dados.
- **Pickle:** é um módulo que implementa protocolos binários para serializar e desserializar uma estrutura de objeto Python. “Pickling” é o processo pelo qual uma hierarquia de objetos Python é convertida em um fluxo de bytes, e “unpickling” é a operação inversa, em que um fluxo de bytes (de um arquivo binário ou objeto semelhante a bytes) é convertido de volta em uma hierarquia de objetos.

Q

R

S

- **SaaS:** Software as a Service ou Software como Serviço é uma modalidade de serviço de cloud se baseia no fato do usuário não precisar se preocupar com a sua instalação, manutenção ou atualização, apenas acessa o sistema pelo seu endereço na internet.
- **Series:** é um array unidimensional, uma lista de valores. Toda Series possui um índice, o index, que dá rótulos a cada elemento da lista.
- **S&OP (Sales & Operation Planning, Planejamento de Operações e Vendas):** é um processo integrado para gerir as áreas de e negócios. A evolução dessa modalidade aconteceu quando as indústrias decidiram integrar as áreas administrativas à produção.
Nas grandes indústrias, é necessário planejar o volume de produtos da melhor maneira possível. Isso porque o volume total é compartilhado em áreas distintas (Comercial, Marketing, Desenvolvimento, Fabricação, Sourcing) e nem todos os sistemas são integrados entre si.
- **Supply Chain Management (Gerenciamento da Cadeia de Suprimentos):** são todas as atividades de compra dos insumos ou produtos, transporte, armazenamento, transformação, embalagem, gerenciamento interno, venda e distribuição aos clientes. Adicione a todas essas etapas a infraestrutura física e de processos necessária para dar suporte às operações.
- **Sell In:** são as vendas do fabricante, da indústria, para o canal – distribuidor ou varejista.
- **Sell Out:** está ligado ao sell-in, é o passo seguinte da venda, a entrega ao cliente final.

- **Serialização:** é o processo de tradução de estruturas de dados ou estado de objeto em um formato que possa ser armazenado e reconstruído posteriormente no mesmo ou em outro ambiente computacional.

T

- **Taxonomia:** ciência ou técnica de classificação.
- **Trade Allowance (TA):** Desconto de negociação ou reduções de preços concedidas a intermediários, como varejistas, para incentivá-los a estocar os produtos de uma organização.

U

V

W

X

Y

Z

1.2 Fontes

A

B

C

D

- **Dask:** <https://docs.dask.org/en/latest>

E

F

G

H

I

- **Index-a-Dora:** <https://indexadora.wordpress.com/category/organizacao-da-informacao>

J

K

L

M

- **Matplotlib:** <https://matplotlib.org>

N

- **Numpy:** <https://numpy.org>

O

- **Open research software:**
<https://openresearchsoftware.metajnl.com/articles/10.5334/jors.148>

P

- **Pandas:** <https://pandas.pydata.org>
- **Python:** <https://www.python.org>
- **Pyplan:** <https://pyplan.org>
- **Pyplan Documentação:** <https://docs.pyplan.org>
- **PyPlan Comunidade:** <https://community.pyplan.org>

Q

R

- **Readthedocs:** <https://readthedocs.com>

S

- **Seabron:** <https://seaborn.pydata.org>
- **Sphinx:** <https://www.sphinx-doc.org>

T

- **Taxonomy:** <https://www.taxonomybootcamp.com/2015/Monday.aspx>

U

V

W

X

- **Xaaray:** <http://xarray.pydata.org>

Y

Z

Conclusão

Entendemos que este é um documento vivo e que não apresenta um estado único, assim este está sempre em constante mudança de acordo o direcionamento de novas metodologias, tecnologias, abordagens de mercado, e transformações no cenário macro e micro econômico.