

Utilização De Um Data Warehouse Para Análise Histórica Do Volume Pluviométrico Do Município De Belo Horizonte No Período De 2010 a 2014.

Use Of A Data Warehouse For Discounts From Historical Analysis Rainfall
From Belo Horizonte From City In Period From 2010 to 2014.

El uso de un depósito de datos para los descuentos a partir del análisis histórico de lluvia de Belo Horizonte a partir de la ciudad En el período de 2010 a 2014.

Fernanda Souza Rezende¹

Diego de Oliveira Souza²

Rodrigo Vitorino Moravia³

RESUMO: A gerência e o uso de recursos hídricos têm sido foco de atenção no Brasil atualmente. O CPTEC/INPE disponibiliza uma extensa base de dados contendo dados climáticos de diversas regiões do país, coletados de diversas formas. Este trabalho tem por objetivo apresentar estes dados em formatos gráficos a fim de possibilitar a comparação de volume pluviométrico no município de Belo Horizonte no período de 2010 a 2014.

Palavras-chave: Dados Públicos, Pluviometria, Precipitação.

Abstract: *The management and use of water resources have been the focus of attention in Brazil today. CPTEC / INPE offers an extensive database of weather data from different parts of the country, collected in several ways. This work aims to present these data in forms of graph to enable the rainfall volume comparison in the city of Belo Horizonte in the period 2010 to 2014.*

Keywords: Public Data, Pluviometry, Precipitation.

Resumen: La gestión y el uso de los recursos hídricos han sido el foco de atención en el Brasil de hoy. CPTEC / INPE ofrece una extensa base de datos de los datos de tiempo en diferentes partes del país, recogidos de varias maneras. Este trabajo tiene como objetivo presentar estos datos en forma de gráfica para permitir la comparación del volumen de lluvias en la ciudad de Belo Horizonte, en el período 2010-2014.

Palabras clave: Datos Públicos, Pluviometría, Precipitación.

1. INTRODUÇÃO

O maior bem da humanidade e todos os seres vivos, está na presença da água. Este recurso que aparentemente é abundante mas tem sido muito mal utilizado, com uma série de desperdícios, poluição e má distribuição.

Atualmente, no Brasil, a utilização de recursos hídricos e controle de reservas de água potável têm sido foco de polêmicas e considerações de muitos estudos. Da mesma forma, no município de Belo Horizonte apresenta diversos problemas que estão relacionados ao índice pluviométrico, poluição

de ribeirões, entupimento de “bocas-de-lobo” que causam transtornos e alagamentos, aterramento de nascentes, asfaltamento de ruas, dentre outros.

O CPTEC/INPE (Centro de Previsão do Tempo e Estudos Climáticos/Instituto Nacional de Pesquisas Espaciais) monitora diariamente dados meteorológicos de diversas regiões do Brasil, através de satélites, boias, balões e plataformas de coletas de dados (PCD). Dentre estes dados, encontram-se os índices pluviométricos por região, umidade relativa do ar, temperatura máxima, média e mínima observados em intervalos de 3 horas.

A PCD instalada na cidade de Belo Horizonte disponibiliza dados de precipitação no município em intervalos regulares, o que permite um acompanhamento do histórico de precipitação na cidade.

Estas informações podem ser recuperadas em formato de planilhas ou arquivos texto, o que torna mais difícil a interpretação e visualização dos dados.

De posse dos dados disponibilizados pelo CPTEC/INPE é possível traçar um perfil de precipitação por região e utilizar estes dados para fornecer informação para auxiliar a tomada de decisão quanto ao uso e racionamento de recursos hídricos.

Desta forma tem sido noticiado uma necessidade constante de racionamento e economia de água por parte das autoridades públicas. Mas será que estas medidas de prevenção já não deveriam ocorrer há algum tempo, já que os dados estão disponibilizados publicamente e historicamente pelo CPTEC/INPE?

Levando em consideração os pontos descritos no problema e na introdução, tem por objetivo utilizar os dados fornecidos pelo CPTEC/INPE para gerar um conjunto de informações mais legível e que possa servir de base para tomada de decisão no tocante a utilização de recursos hídricos no município de Belo Horizonte. Com isso, coletaremos os dados que serão armazenados em um banco de dados, onde serão realizados os devidos tratamentos bem como seu enriquecimento, como por exemplo de estações do ano, nos permitindo comparar através de análises gráficas fazendo uso da ferramenta *Excel*, possibilitando assim analisarmos historicamente informações históricas do período de 2010 a 2014.

1.1. Estrutura do Trabalho

Este trabalho encontra-se organizado em 6 capítulos.

O capítulo 1, apresenta a introdução ao assunto discutido no trabalho e a organização do trabalho em si. No capítulo 2, este trabalho apresenta os conceitos de Business Intelligence (BI) utilizados para realizar a coleta, análise, tratamento e apresentação dos dados. O capítulo 3 apresenta os dados coletados e as ferramentas utilizadas para a confecção deste trabalho. No capítulo 4, apresentam-se e discutem-se brevemente os resultados obtidos e as considerações finais deste trabalho. Por fim, o capítulo 5 apresenta as referências bibliográficas utilizadas.

1.2. Metodologia

Este trabalho foi dividido em quatro etapas: coleta, tratamento, análise e escrita.

A etapa de coleta é composta por extração dos dados em formatos CSV do site do CPTEC/INPE e migração dos formatos de dados para xls.

Já na etapa de tratamento, os dados coletados foram normalizados e inseridos na ferramenta *Excel* onde foram tratados e realizou-se a geração dos gráficos que foram utilizados como insumo para etapa de análise dos dados, na qual os gráficos gerados foram analisados com o intuito de se identificar padrões.

Por fim, na etapa de escrita, confeccionou-se este documento.

2. REFERENCIAL TEÓRICO

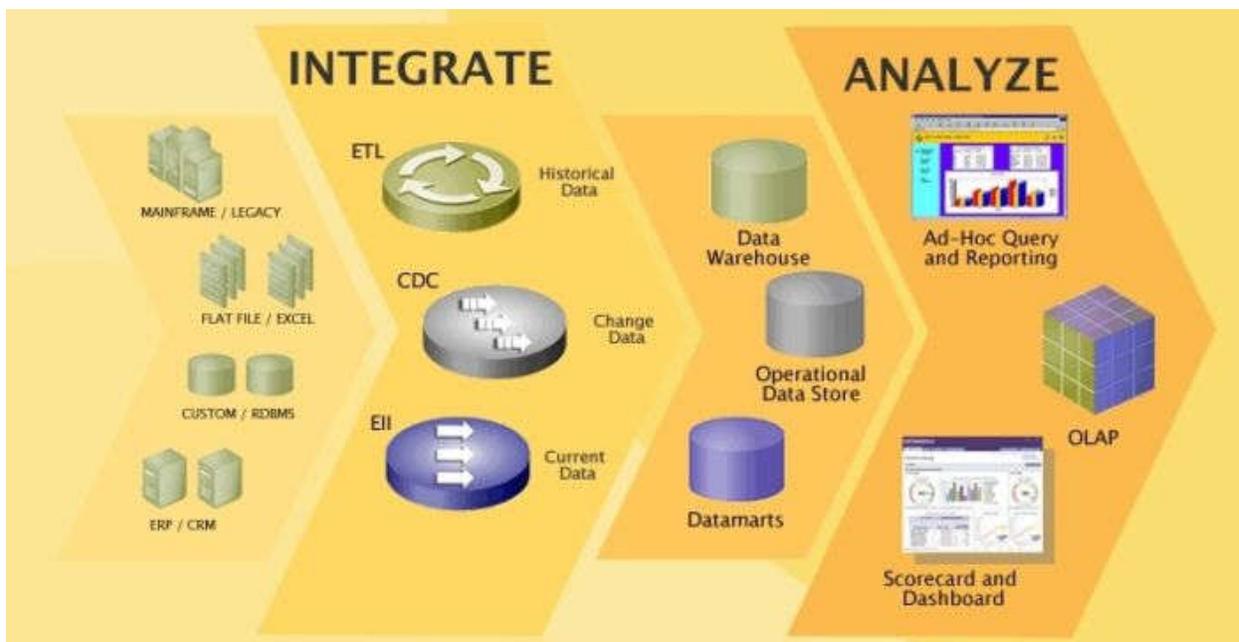
2.1. Business Intelligence

Business Intelligence (BI) pode ser definido de forma resumida como o uso de diversas fontes de dados para embasamento da tomada de decisões

relativas a o negócio de uma empresa, ou seja, utilizar os recursos de dados disponíveis para que sejam tomadas melhores decisões. Isso envolve acesso a dados, análise e o descobrimento de novas oportunidades(Barbieri 2001, Almeida 2009).

Dito isso, um processo de BI pode ser definido como a extração de dados, consolidação e tratamento dos mesmos e apresentá-los de forma a auxiliar a tomada de decisão. A figura 1 apresenta uma arquitetura de sistemas de Business Intelligence.

Figura 1 - Arquitetura de sistemas de Business Intelligence



Fonte: <http://www.dashboardinsight.com/>

Como mostrado na figura 1 um processo de BI é composto por um conjunto de técnicas, das quais as seguintes se mostram mais relevantes para o âmbito deste trabalho:

- *Data Warehousing*: armazém de dados, composto por dados obtidos de diversas origem através de processos de ETL (*Extract, Transform and Load*). Este armazém utiliza uma modelagem dimensional de banco de dados.

- *Online Analytical Processing* (OLAP): Proporciona a utilização e manipulação de dados armazenados em *Data Warehouses* com velocidade, facilidade e segurança.

2.2. Data Warehouse

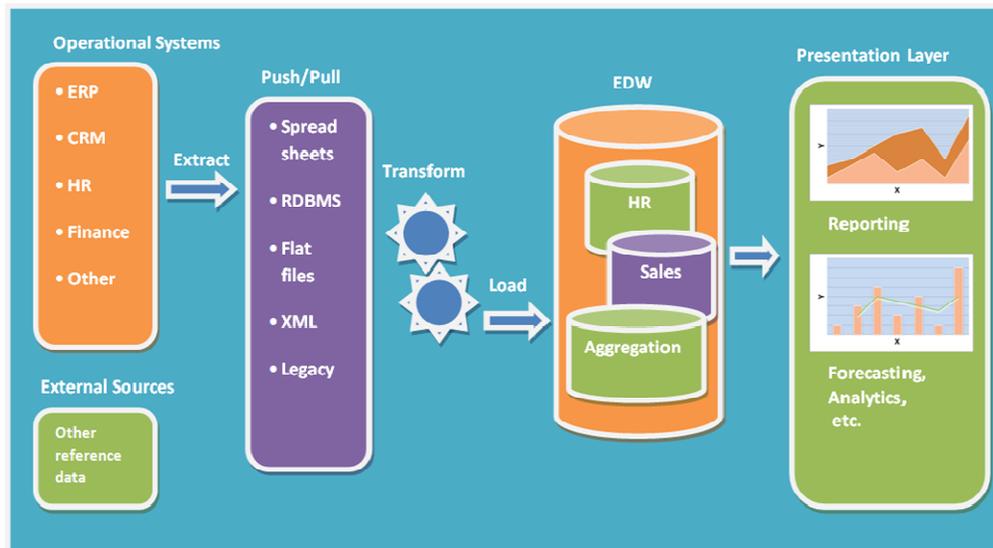
Data Warehouse (DW) é um conjunto de dados organizados por assunto, integrados com variação baseada no tempo e não volátil (INMON 1992).

Os DWs são construídos para satisfazer as necessidades de gestores, e não de usuários comuns de sistemas. Em um DW a informação é limpa, organizada e consistente e se encontra organizada de forma a permitir a fácil interpretação por um gestor de negócio. Além disso, diferente de outros sistemas que mantêm apenas dados atuais, num DW encontram-se tanto as informações atuais quanto históricas (Almeida, 1999).

Em resumo, em um DW encontram-se dados provenientes de diversas fontes (planilhas, documentos, bancos de dados de sistemas computacionais, etc.) e estes dados são reorganizados de forma a preencherem um banco de dados dimensional organizado por assunto. Ao processo de coleta, tratamento e inserção destes dados num DW, dá-se o nome de ETL (*Extraction, Transformation, Load*).

O processo de ETL respectivamente em: extrair dados de diversas origens, realizar as transformações necessárias nestes dados e por fim carrega-los em um DW. A figura 2, ilustra o processo de ETL e sua interação com o processo de BI como todo.

Figura 2 - Processo de ETL



Fonte: <http://www.commuters.eu/>

Como mostrado na figura 2, os dados armazenados em um DW podem ser utilizados para apresentação ao usuário de diversas formas com o auxílio de diversas ferramentas disponíveis no mercado.

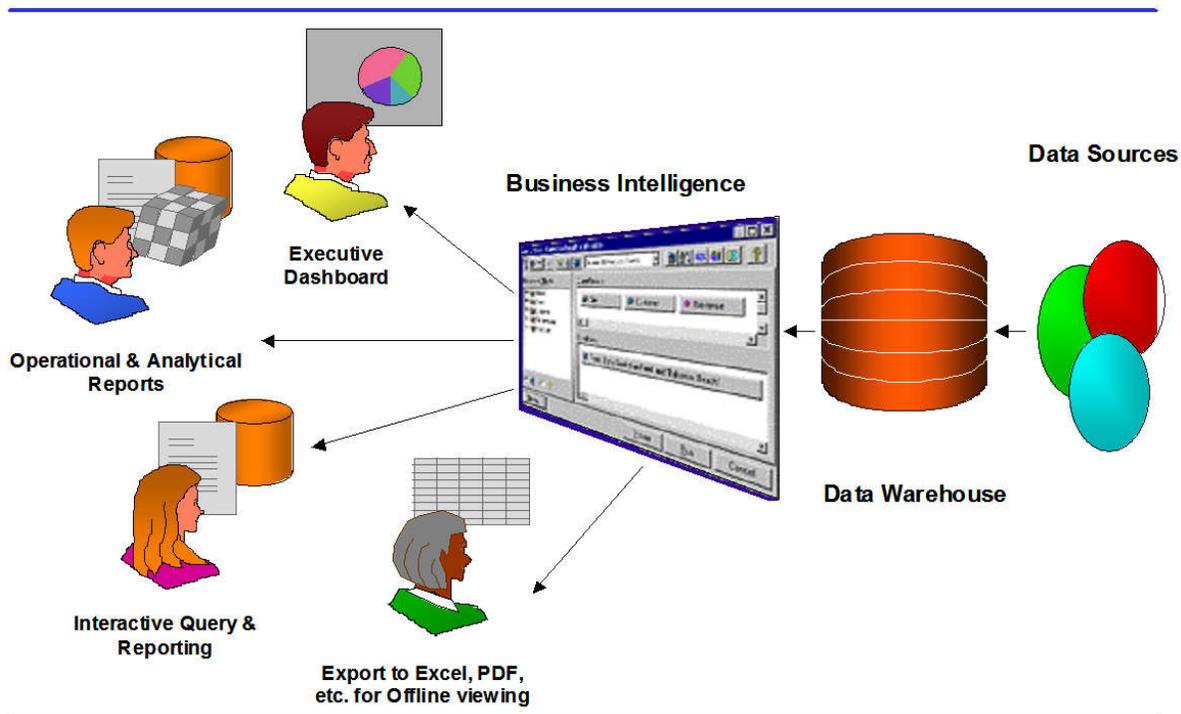
2.3. Online Analytical Processing

Uma definição simples para OLAP (*Online Analytical Processing*) seria dizer que é a tecnologia por trás das muitas ferramentas de BI existentes, uma ferramenta poderosa para descobrimento de dados que inclui capacidade para geração de relatórios, cálculos complexos e análise preditiva de cenários.

O OLAP com as características descritas acima permite ao usuário final analisar dados em diversas formas e dimensões, dando assim a visão e o entendimento necessário para a melhoria na tomada de decisões (OLAP, 2015).

A figura 3 mostra a utilização do OLAP e a visualização das facilidades provenientes de um sistema OLAP por usuários finais.

Figura 3 - Uso de vantagens do OLAP por usuários finais



2.4. Índice Pluviométrico

Segundo Santomauro (2015), “o índice pluviométrico refere-se a quantidade de chuva por metro quadrado em determinado local e em determinado período”. Este índice é calculado em milímetros pelo aparelho conhecido como pluviômetro (figura 4) por várias estações meteorológicas espalhadas no país. É encontrado vários modelos diferentes, o aparelho é basicamente um funil que recebe a água e faz o registro através de movimentos que duas básculas, instaladas no instrumento, se movimenta, gerando assim sinais elétricos que são enviados para as estações.

Figura 4 – Pluviômetro



Fonte: <http://revistaescola.abril.com.br/>

3. DADOS PLUVIOMÉTRICOS DE BELO HORIZONTE

O CPTEC/INPE disponibiliza o SINDA (Sistema Integrado de Dados Ambientais) para consulta de informações relacionadas a estudos climáticos em diversas regiões do Brasil.

Os dados disponibilizados são coletados através de PCDs (Plataformas de coletas de dados), também chamadas de Estações Ambientais Automáticas. Estas PCDs surgiram da necessidade de se coletar informações meteorológicas das regiões brasileiras para serem utilizadas na previsão do tempo, porém também são utilizadas por empresas como hidrelétricas para controlar níveis de suas barragens e tomar decisões como a de abrir ou não as comportas das usinas. Em geral as PCDs, são munidas de células solares e baterias o que torna sua vida útil quase indefinida. A comunicação com as PCDs é feita via satélite o que permite que os dados coletados sejam acessados por usuários de diversas formas em diversos lugares do planeta.

Na página WEB do SINDA é possível realizar a extração de dados históricos de cada uma das PCDs instaladas pelo Brasil. Neste trabalho extraiu-se os dados da PCD instalada no município de Belo Horizonte (ID 32513) para todos os meses do ano, nos anos de 2010 a 2014.

Os dados foram extraídos em formatos xls, em 5 planilhas separadas por ano. Cada planilha seguia o formato de dados apresentado na figura 4. Para que os dados pudessem ser inseridos posteriormente em um banco de dados removeu-se o cabeçalho de todas as planilhas e apenas uma planilha foi gerada.

Figura 5 – Formato da planilha

DataHora	Bateria (Volts)	CorrPSol (Logico)	DirVelVentoMax (oNV)	DirVento (oNV)	Pluvio (mm)	RadSolAcum (MJ/m²)	TempAr (oC)	TempMax (oC)	TempMin (oC)	UmidRel (%)	VelVento3m (m/s)	VelVentoMax (m/s)
14/04/2015 21:00	12.2	0	340	10	6	2	21	25	17.5	55	3.1	6
14/04/2015 18:00	12.9	1	290	290	6	6.6	24	25.5	17.5	39	2.8	6.9
14/04/2015 15:00	12.9	1	350	350	6	8.3	24	26	17.5	43	2.2	6.5
14/04/2015 12:00	12.9	1	360	50	6	2.9	21.5	26	17.5	57	2	5.5
14/04/2015 09:00	12.1	1	30	30	6	0	17.5	26	16.5	67	1.6	4.9
14/04/2015 06:00	12.1	0	10	20	6	0	18	26	16	62	1.6	4.8
14/04/2015 03:00	12.2	0	350	20	6	0	19	26	16	101	1.8	6.6
14/04/2015 00:00	12.2	0	340	0	6	0	20.5	26	16	64	1.9	6.7
13/04/2015 21:00	12.2	0	350	210	6	1.4	22.5	26	16	51	2.4	5.9
13/04/2015 18:00	12.9	1	350	310	6	7.2	25	26	16	38	2.5	6.5
13/04/2015 15:00	12.9	1	340	10	6	8.6	24	24.5	16	42	2.4	7.1
13/04/2015 12:00	12.9	1	60	0	6	2.9	21.5	24.5	16	57	2.7	6.9
13/04/2015 09:00	12.1	1	60	290	6	0	16.5	24.5	16	81	1.4	3
13/04/2015 06:00	12.1	0	40	50	6	0	17.5	24.5	16.5	75	2.2	3
13/04/2015 03:00	12.1	0	60	350	6	0	18.5	24.5	16.5	71	0.6	4.4
13/04/2015 00:00	12.2	0	300	10	6	0	20.5	24.5	16.5	62	2.4	5.6
12/04/2015 21:00	12.4	0	340	320	6	4.5	24	24.5	16.5	42	1.2	4.2
12/04/2015 18:00	12.4	1	340	320	6	4.5	24	24.5	16.5	42	1.2	4.2
12/04/2015 15:00	12.9	1	350	260	6	8.7	23	24	16.5	47	1.7	5.7
12/04/2015 12:00	12.1	1	50	40	6	0	16.5	24	16.5	81	2	3.5

Fonte: SINDA

Por fim, após esses ajustes, gerou-se uma única planilha, contendo os dados de todos os anos de 2010 a 2014 que foi usado para realização de uma carga dos dados através de uma ferramenta Import/Export que acompanha o banco de dados SQL Server da empresa Microsoft. Alguns tratamentos foram realizados para inclusão, como por exemplo, a coluna que trata da Estação do Ano. Na seção seguinte, são apresentados os gráficos gerados com dados acessados no banco de dados separados por ano, e uma breve análise foi realizada e posteriormente uma análise comparativa dos 5 anos.

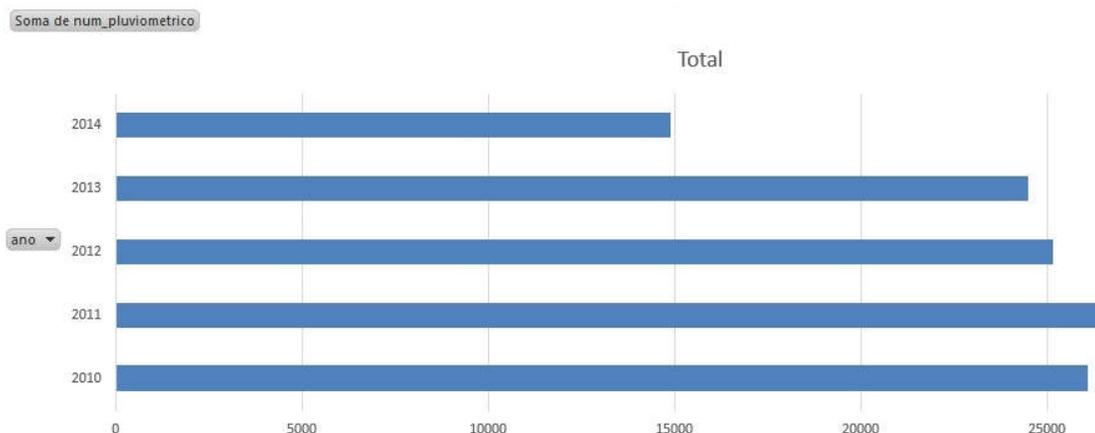
4. ANÁLISE DOS RESULTADOS

Após o tratamento dos dados, utilizou-se o *Microsoft Excel*. Com os dados inseridos na plataforma, gerando 5 gráficos, um por mês para cada um dos 5 anos do período selecionado para análise. Nesta seção estão apresentados cada ano separadamente com o detalhamento da precipitação durante os anos.

O ano de 2010 foi escolhido como modelo, e para o mesmo todos os meses serão apresentados, para os anos seguintes (2011 a 2014) serão apresentados apenas os meses que foram julgados relevantes para o estudo.

O Gráfico 1 apresenta a soma total do índice pluviométrico acumulado dos últimos anos. Percebe-se uma ligeira queda nos últimos três anos, exceto no último que apresentou uma que acentuada neste volume.

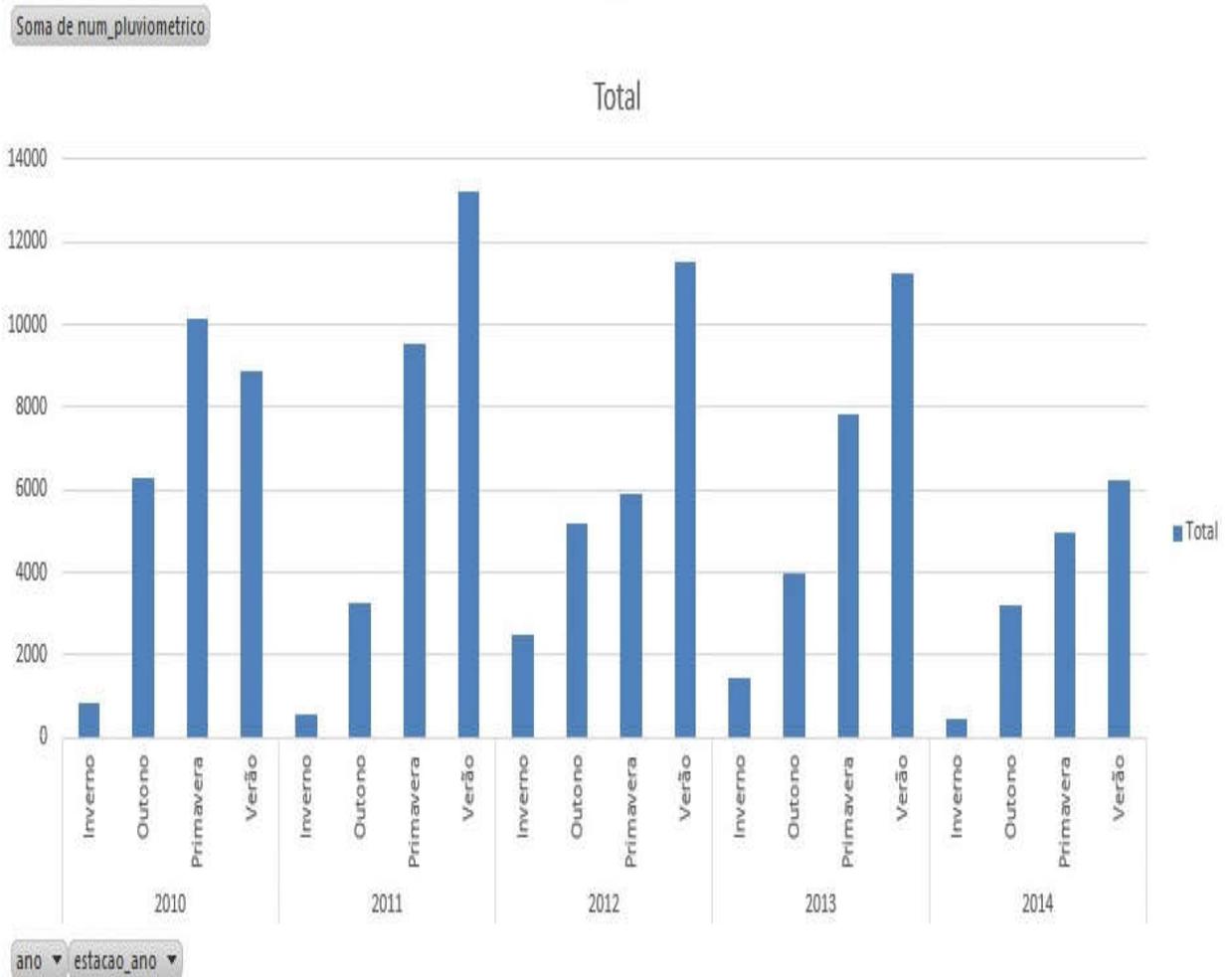
Gráfico1: Índice pluviométrico calculado por ano



Fonte: Dados da pesquisa

Buscando entender qual o período do ano que provocou a queda do índice pluviométrico, realizou-se a análise por estações, nos últimos anos, conforme apresentado no Gráfico 2. Percebe-se uma queda em todas as estações no decorrer do ano de 2014, principalmente no verão que sofreu uma queda bastante acentuada. Provavelmente este volume baixo refletiu-se nos dados que devem ter sido gerados nos primeiros meses do ano de 2015 (dados estes que não fazem parte deste trabalho), como vivenciamos e conseqüentemente gerou vários reflexos e campanhas educativas para o uso correto da água.

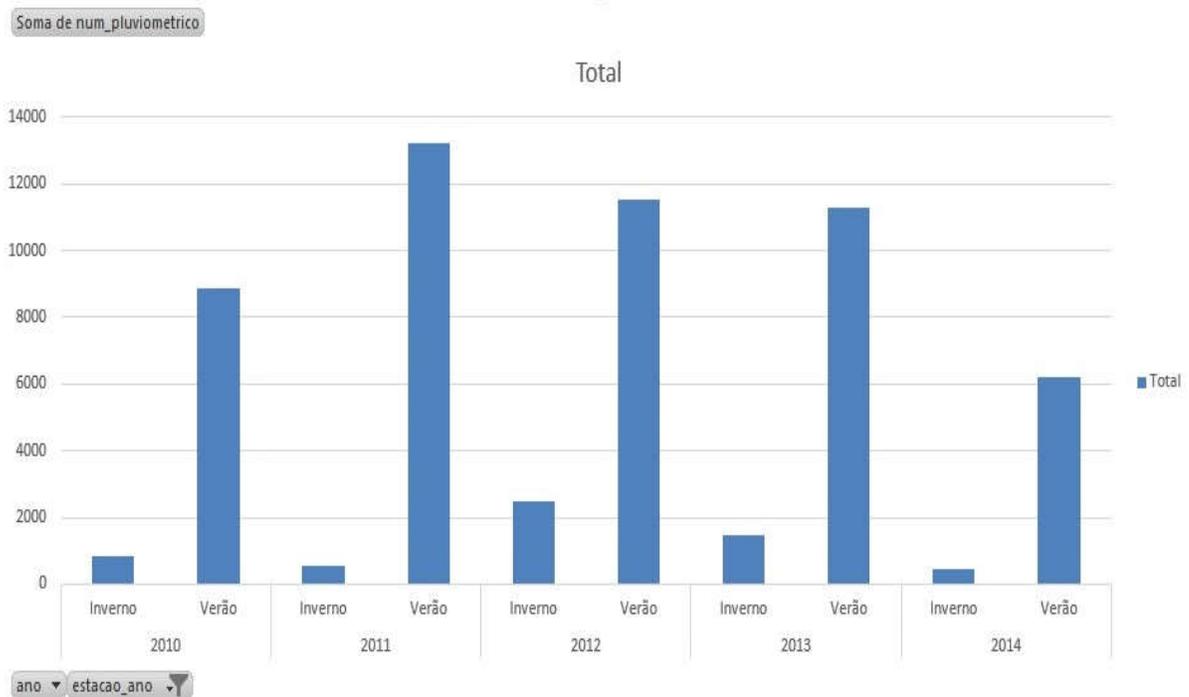
Gráfico2: Índice pluviométrico calculado por estações por ano



Fonte: Dados da pesquisa

O Gráfico 3 comprova também que o inverno historicamente é uma estação do ano com um índice pluviométrico muito fraco, e se não chover o suficiente no Verão, os reflexos se estenderão para o restante do ano.

Gráfico3: Estações que possuem maior e menor índice de chuva

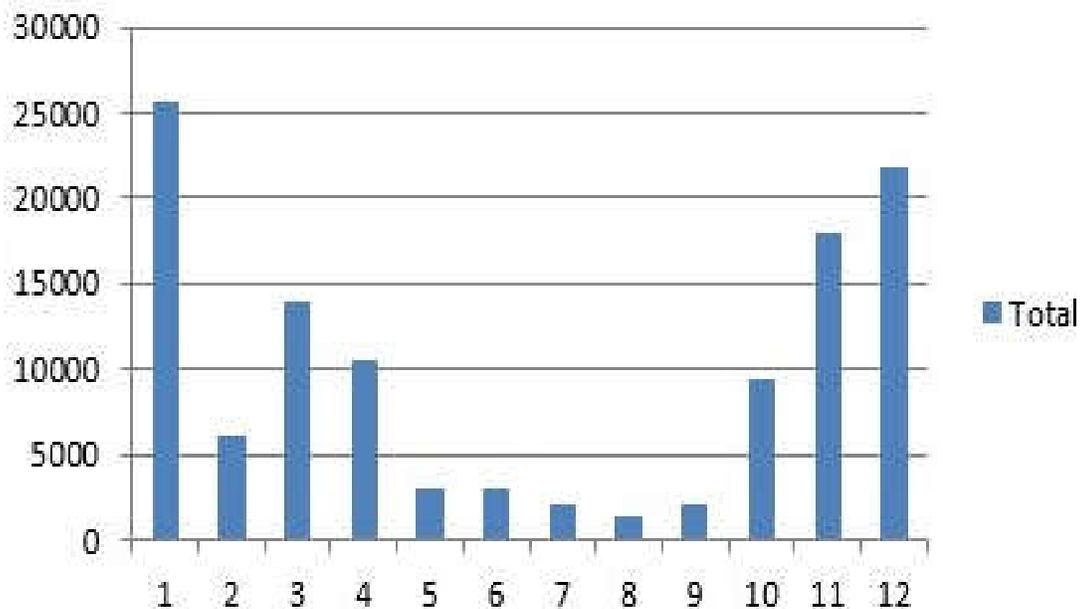


Fonte: Dados da pesquisa

O Gráfico 4 apresenta o acumulado dos meses de todos os anos, ou seja, o total de chuva acumulada nos 5 anos para o mês de setembro e assim sucessivamente. Percebe-se que o mês de janeiro, representado no gráfico pelo número 1, possui o maior volume de chuva. Já para os meses de maio até setembro, respectivamente representados no gráfico pelos números 5 a 9, são os meses que apresentam o menor volume. Estes meses representam as estações do outono e inverno, como já visto em outra análise

Gráfico4: Acúmulo de chuva de acordo com os meses

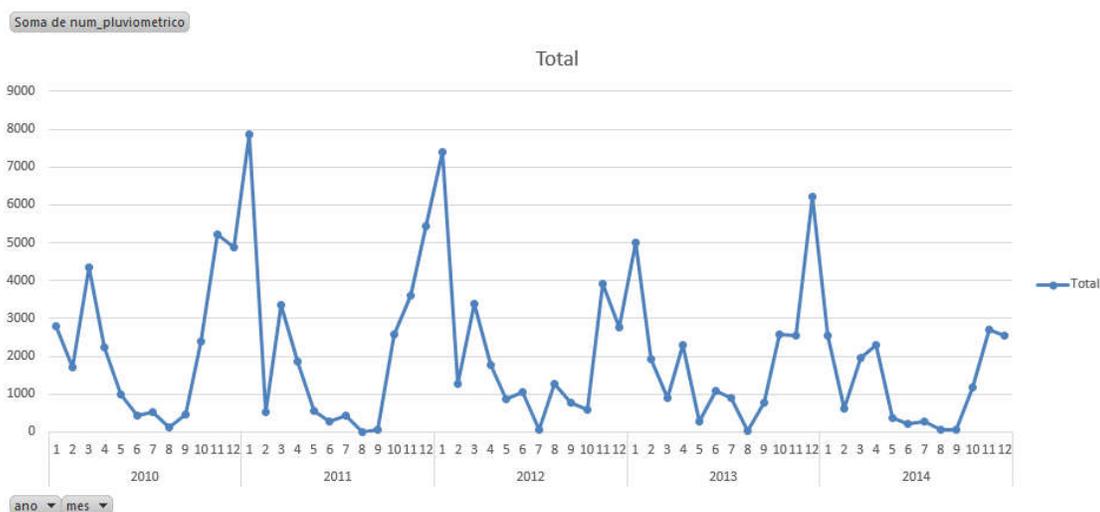
Total



Fonte: Dados da pesquisa

Conforme Inmon (1992), oGráfico 5 mostra a evolução histórica do índice pluviométrico mensal de cada um dos anos, onde percebe-se que a curva deste gráfico vai diminuindo com o passar do tempo, ou seja, o ano de 2010 teve mais chuvas que o de 2011 e assim por diante. Chegando então ao ano de 2014 percebe-se uma nítida diferença em comparação ao ano de 2010, o que vem a comprovar a diminuição que tem ocorrido ao longo dos anos.

Gráfico5: Acúmulo de chuva de acordo com os anos separados por mês



Fonte: Dados da pesquisa

5. CONCLUSÃO

Observamos então que o nível do índice pluviométrico abaixou drasticamente do ano de 2013 para o ano de 2014, mas este mesmo índice já vinha sofrendo quedas desde 2010 para 2011. Atualmente governo estadual tem feito medidas educativas para que a população tenha consciência e use a água com mais atenção e menos desperdício. De acordo com as pesquisas e dos dados aqui apresentados, dados estes que são públicos, o governo anterior já poderia ter tomado medidas mais sérias, mesmo antes que chegasse em um nível tão preocupante. No final do ano de 2012 e início de 2013, de acordo com as análises apresentadas, mais especificamente o gráfico 5, a queda foi bem acentuada. Este seria um bom momento para começar as campanhas e ações de melhorias vinda do governo.

Desta forma podemos dizer que os objetivos propostos pelo trabalho foram alcançados, valendo ressaltar que esses dados coletados são de domínio público, ou seja, todos tem acesso. Mas de nada adianta se não é observado e ações não são tomadas por parte dos governantes e de uma conscientização da população.

No trabalho a maior dificuldade foi na extração e limpeza dos dados, pois os dados são coletados mais de uma vez por dia e nem sempre seguia o mesmo padrão. Foi necessário conhecer também os recursos da ferramenta Excel para importação dos dados e geração dos gráficos.

Como sugestão para trabalho futuro seria interessante fazer uma comparação com o ano de 2015 ou mesmo de anos anteriores além de apresentar os resultados obtidos com as campanhas do governo.

REFERÊNCIAS

ALMEIDA, Sueli. **Getting Started with Data Warehouse and Business Intelligence**. San Jose, Califórnia: IBM Corporation, 1999.244p.

BARBIERI, Carlos. **BI – Business Intelligence: Modelagem e Tecnologia**. Rio de Janeiro, Axcel Books do Brasil Editora, 2001.424p.

INMON, W.H. **Bulding the Data Warehouse**. John Wiley & Sons Inc. 1992, 1ª edição

OLAP.Com. **What is OLAP? Definition of OLAP, Advantges and Uses**. 2015. Disponível em: <<http://olap.com/olap-definition/>>. Acesso em 22 abr. 2015.

Quadro, M.F.L; Machado, L.H.R; Calbete, S,; Batista, N.N.M; Oliveira, G.S. **Climatologia da Precipitação e Temperatura**. CPTEC/INPE. Disponível em: <<http://climanalise.cptec.inpe.br/~rclimanl/boletim/cliesp10a/chuesp.html>>. Acesso em 25 abr. 2015

SANTOMAURO, Beatriz. **O que é e como se calcula o índice pluviométrico?** Disponível em: <<http://revistaescola.abril.com.br/geografia/fundamentos/como-se-calcula-indice-pluviometrico-476502.shtml>>. Acesso em 02 out. 2015.