

Hoje a gente vai começar uma série nova aqui no canal de análise de dados e o que é análise de dados bons talvez você conhece essa área, talvez não, mas de maneira bem simples ela é nada mais nada menos do que você olhar para um conjunto de dados e tentar extrair informações dele ele é um campo multidisciplinar que envolve conhecimento de Ciências da Computação Inteligência Artificial matemática e estatística entre outros e por que esse campo tem sido tão falado estouro tão importante nos dias de hoje a gente vai começar uma série nova aqui no canal de análise de dados e o que é análise de dados bons talvez você conhece essa área, talvez não, mas de maneira bem simples ela é nada mais nada menos do que você olhar para um conjunto de dados e tentar extrair informações dele ele é um campo multidisciplinar que envolve conhecimento de Ciências da Computação Inteligência Artificial matemática e estatística entre outros e por que esse campo tem sido tão falado estouro tão importante nos dias de hoje porque é um momento em que a gente está vivendo em que existe muita informação sendo coletada principalmente Esses aparelhos aqui que são os nossos smartphones toda a vez que a gente está usando smartphone que a gente está usando no aplicativo provavelmente existem dados sendo coletados do nosso uso e esses dados eles são utilizados para, por exemplo criar sistemas de recomendação de produtos de serviços para gente para analisar e criar serviços que nos indica em quanto tempo a gente está passando no celular para que a gente possa gerenciar e não causar danos a nossa própria saúde enfim existe uma série de aplicações nos dias de hoje que conemuitos dados e por causa disso esse campo da análise de dados começou a ganhar tanta relevância e bom Eu particularmente gosto muito dessa área eu vou especializar nisso eu fiz um mestrado Justamente na análise na modelagem de dados Eu gosto muito de conhecer são muito feliz em começar essa série aqui no canal dividiu um pouco dos meus conhecimentos sobre isso com vocês bom então nada que a gente já tem agora uma noção geral sobre o que é análise de dados é legal a gente conversar discutir o que é um dado de fato porque às vezes a gente confunde quando a gente tá! falando com localmente corriqueiramente dado com informação e eles são conceitos um pouco distintos então dado ele é um registro de eventos de fatos que ocorrem ao nosso redor está um pouco abstrato, mas dá uma olhada aqui nessa coluna essa coluna A gente tem uma tabela com números a princípio a gente não sabe exatamente o que é isso e a gente não sabe porque a gente não tem a informação do que são esses da informação é justamente o que organiza o que para sentido aquele dado então um dado sem informação ele não tem muito sentido para gente então, por exemplo olhando para essa mesma coluna agora que eu te digo que esses são dados de temperatura ao longo da semana passada na Cidade do Rio de Janeiro eu estou tirando informação sobre o que eram aqueles números brutos e você consegue entender bom esse dado é sobre isso tudo de variação basicamente é essa dado é o bruto é o evento é o registro isolado informação é a camada de sentido que a gente dá o dado que geralmente a gente faz análise justamente para conseguir mais informações sobre algum campo do conhecimento então, por exemplo que ao invés de ter essa coluna aqui da temperatura se eu tivesse uma tabela maior com registro dos últimos 365 dias ou nos últimos cinco anos da temperatura diária ou da médica de temperatura diária na Cidade do Rio de Janeiro eu tenho mais curiosidade de dados maior que eu poderia utilizar para analisar como foi a variação de temperatura no passado e inclusive podemos utilizar para tentar prever informações sobre o futuro e aí É clara essa previsão ela vai ter uma imagem de erro ela pode ter ou não vier esse tempo técnico, mas a gente vai discutir mais para frente, mas o fato é que eu consigo prever uma possível temperatura se ela vai concordar com a realidade se ela vai ter 100% 80% de acerto de chance de estar certa isso é uma outra história já que os dados eles são registros de

eventos e de fato socorridos é importante para que a gente possa fazer uma análise de dados, na verdade é fundamental porque a gente tem então uma fonte de dados que a gente saiba que dados são eles que a gente quer analisar porque falar de dados em geral como eu falei muito para gente que são colhidos o tempo inteiro se separa alguns sites para a gente começar a olhar então dá uma olhada nesse primeiro DATAHUB.IO/COLLECTIONS Iha que bacana ele tem várias coleções que ele chama aqui conjunto de dados então sobre aquele clima, educação, futebol americano, saúde, dados de inflação, tudo sobre os Estados Unidos e olha esse outro site aqui DATA.WORLDBANK.ORG é porque o site do Banco Mundial ele também disponibiliza dados isso é muito legal para quem não sabe eu gosto muito de desenvolvimento de países de projetos diferentes em paizão ao redor do mundo inteiro aqui você vai encontrar dados sociais que você possa analisar tem esse outro site aqui que é da NASA muito legal também disponibilizam algum conjunto de dados sem o site aqui do Governo dos Estados Unidos e tem o site que eu acho que é muito legal isso também que existe eu não sei se você já sabe que é um site de dados abertos do Brasil o governo brasileiro não tem um portal de dados abertos isso é muito legal porque a gente pode olhar para dados de políticas públicas, por exemplo aqui do Brasil Então olha só aqui a gente Abre o site e a gente vê onde eles A gente tem base cartográfica micro dados de senso de educação dados de orçamento tem uma série de outros dados bacanas frente a esse tempo de site que tem vários conjuntos usados diferentes resolvi escolher justamente o portal de dados brasileiros a gente tirar o primeiro data 7 que a gente vai analisar Então a gente vai analisar os micro dados do Exame Nacional de Ensino Médio e olha só vou-te dar uma amostra já porque a gente vai poder fazer então esse aqui ele é um dicionário do conjunto de dados ele é justamente a informação sobre o dado lembra que eu expliquei lá no início diferença de informação e dado então aqui eu tenho informação sobre o que tem no meu conjunto de dados se eu não tivesse esse dicionário seria muito mais difícil para eu começar a analisar o conjunto de dados porque ia ter um monte de colunas um conjunto de dados provavelmente muito extenso e eu não ia saber o que é cada informação e aqui justamente ele diz-me o nome das variáveis, ou seja o nome das colunas da tabela e que ela significa então tem o número de inscrição do candidato um ano que ele fez o Enem eu tenho dados de onde ele mora, sexo, estado civil aqui gente é importante a gente falar que esse tipo de dados que fica público ele não é uma invasão a privacidade das pessoas porque ele é pesado Eles não têm, por exemplo um nome completo da pessoa o CPF enfim esses dados protegem geralmente a informação sensível aí que a gente tem só o que a gente pode de fato olhar, mas vamos continuar vendo olha que legal a gente tem estado civil o ano de conclusão no ensino médio a gente tem dados como se a pessoa possui algum tipo de deficiência auditiva deficiência física a gente tem muitos dados legais eu acho que aqui tem uma menina de ouro mesmo porque a gente possa analisar Quem são os alunos que fazem o Enem e através desse conjunto de dados para poder mostrar para vocês conceitos básicos da análise de dados em Python Então vou apresentar as principais pacotes as principais métodos dentro desses pacotes para você fazer uma exploração dos dados e mais para frente também vai fazer visualizações de dados que é uma área que eu acho muito legal então é isso gente nesse primeiro vídeo dá um contexto para vocês dessa nova série A gente viu que a análise de dados Qual a diferença de idade de informação que tipo de dados que a gente vai começar a trabalhar daqui para frente que são Dados abertos no portal de dados abertos brasileiro.oje porque é um momento em que a gente está vivendo em que existe muita informação sendo coletada principalmente Esses aparelhos aqui que são os nossos smartphones toda a vez que a gente está usando smartphone que a gente está usando no aplicativo provavelmente existem dados sendo coletados do nosso

uso e esses dados eles são utilizados para, por exemplo criar sistemas de recomendação de produtos de serviços para gente para analisar e criar serviços que nos indica em quanto tempo a gente está passando no celular para que a gente possa gerenciar e não causar danos a nossa própria saúde enfim existe uma série de aplicações nos dias de hoje que conemuitos dados e por causa disso esse campo da análise de dados começou a ganhar tanta relevância e bom Eu particularmente gosto muito dessa área eu vou especializar nisso eu fiz um mestrado Justamente na análise na modelagem de dados Eu gosto muito de conhecer são muito feliz em começar essa série aqui no canal dividiu um pouco dos meus conhecimentos sobre isso com vocês bom então nada que a gente já tem agora uma noção geral sobre o que é análise de dados é legal a gente conversar discutir o que é um dado de fato porque às vezes a gente confunde quando a gente tá! falando com localmente corriqueiramente dado com informação e eles são conceitos um pouco distintos então dado ele é um registro de eventos de fatos que ocorrem ao nosso redor está um pouco abstrato, mas dá uma olhada aqui nessa coluna essa coluna A gente tem uma tabela com números a princípio a gente não sabe exatamente o que é isso e a gente não sabe porque a gente não tem a informação do que são esses da informação é justamente o que organiza o que para sentido aquele dado então um dado sem informação ele não tem muito sentido para gente então, por exemplo olhando para essa mesma coluna agora que eu te digo que esses são dados de temperatura ao longo da semana passada na Cidade do Rio de Janeiro eu estou tirando informação sobre o que eram aqueles números brutos e você consegue entender bom esse dado é sobre isso tudo de variação basicamente é essa dado é o bruto é o evento é o registro isolado informação é a camada de sentido que a gente dá o dado que geralmente a gente faz análise justamente para conseguir mais informações sobre algum campo do conhecimento então, por exemplo que ao invés de ter essa coluna aqui da temperatura se eu tivesse uma tabela maior com registro dos últimos 365 dias ou nos últimos cinco anos da temperatura diária ou da médica de temperatura diária na Cidade do Rio de Janeiro eu tenho mais curiosidade de dados maior que eu poderia utilizar para analisar como foi a variação de temperatura no passado e inclusive podemos utilizar para tentar prever informações sobre o futuro e aí É clara essa previsão ela vai ter uma imagem de erro ela pode ter ou não vier esse tempo técnico, mas a gente vai discutir mais para frente, mas o fato é que eu consigo prever uma possível temperatura se ela vai concordar com a realidade se ela vai ter 100% 80% de acerto de chance de estar certa isso é uma outra história já que os dados eles são registros de eventos e de fato socorridos é importante para que a gente possa fazer uma análise de dados, na verdade é fundamental porque a gente tem então uma fonte de dados que a gente saiba que dados são eles que a gente quer analisar porque falar de dados em geral como eu falei muito para gente que são colhidos o tempo inteiro se separa alguns sites para a gente começar a olhar então dá uma olhada nesse primeiro DATAHUB.IO/COLLECTIONS Iha que bacana ele tem várias coleções que ele chama aqui conjunto de dados então sobre aquele clima, educação, futebol americano, saúde, dados de inflação, tudo sobre os Estados Unidos e olha esse outro site aqui DATA.WORLDBANK.ORG é porque o site do Banco Mundial ele também disponibiliza dados isso é muito legal para quem não sabe eu gosto muito de desenvolvimento de países de projetos diferentes em paizão ao redor do mundo inteiro aqui você vai encontrar dados sociais que você possa analisar tem esse outro site aqui que é da NASA muito legal também disponibilizam algum conjunto de dados sem o site aqui do Governo dos Estados Unidos e tem o site que eu acho que é muito legal isso também que existe eu não sei se você já sabe que é um site de dados abertos do Brasil o governo brasileiro não tem um portal de dados abertos isso é muito legal porque a gente pode olhar

para dados de políticas públicas, por exemplo aqui do Brasil Então olha só aqui a gente Abre o site e a gente vê onde eles A gente tem base cartográfica micro dados de senso de educação dados de orçamento tem uma série de outros dados bacanas frente a esse tempo de site que tem vários conjuntos usados diferentes resolvi escolher justamente o portal de dados brasileiros a gente tirar o primeiro data 7 que a gente vai analisar Então a gente vai analisar os micro dados do Exame Nacional de Ensino Médio e olha só vou-te dar uma amostra já porque a gente vai poder fazer então esse aqui ele é um dicionário do conjunto de dados ele é justamente a informação sobre o dado lembra que eu expliquei lá no início diferença de informação e dado então aqui eu tenho informação sobre o que tem no meu conjunto de dados se eu não tivesse esse dicionário seria muito mais difícil para eu começar a analisar o conjunto de dados porque ia ter um monte de colunas um conjunto de dados provavelmente muito extenso e eu não ia saber o que é cada informação e aqui justamente ele diz-me o nome das variáveis, ou seja o nome das colunas da tabela e que ela significa então tem o número de inscrição do candidato um ano que ele fez o Enem eu tenho dados de onde ele mora, sexo, estado civil aqui gente é importante a gente falar que esse tipo de dados que fica público ele não é uma invasão a privacidade das pessoas porque ele é pesado Eles não têm, por exemplo um nome completo da pessoa o CPF enfim esses dados protegem geralmente a informação sensível aí que a gente tem só o que a gente pode de fato olhar, mas vamos continuar vendo olha que legal a gente tem estado civil o ano de conclusão no ensino médio a gente tem dados como se a pessoa possui algum tipo de deficiência auditiva deficiência física a gente tem muitos dados legais eu acho que aqui tem uma menina de ouro mesmo porque a gente possa analisar Quem são os alunos que fazem o Enem e através desse conjunto de dados para poder mostrar para vocês conceitos básicos da análise de dados em Python Então vou apresentar as principais pacotes as principais métodos dentro desses pacotes para você fazer uma exploração dos dados e mais para frente também vai fazer visualizações de dados que é uma área que eu acho muito legal então é isso gente nesse primeiro vídeo dá um contexto para vocês dessa nova série A gente viu que a análise de dados Qual a diferença de idade de informação que tipo de dados que a gente vai começar a trabalhar daqui para frente que são Dados abertos no portal de dados abertos brasileiro.