



01

Distribuição pelo Histograma

Transcrição

Continuando o trabalho de análise, você recebeu um novo desafio: Verificar os dados sobre o tempo do serviço de entrega da Jump Cats. Você precisará descobrir se o tempo é regular, se a demora é grande, qual é a velocidade média. A empresa quer ter uma ideia de como o serviço está se comportando.

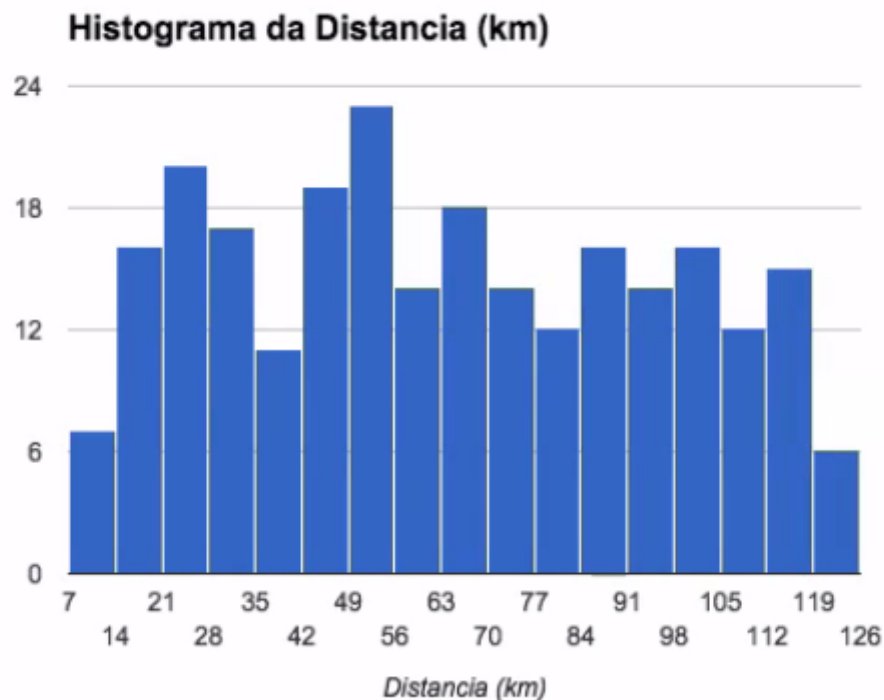
Foi fornecida uma planilha com a distância e o tempo. Nossa tarefa é entender para chegar a algum valor ou estatísticas úteis para a empresa.

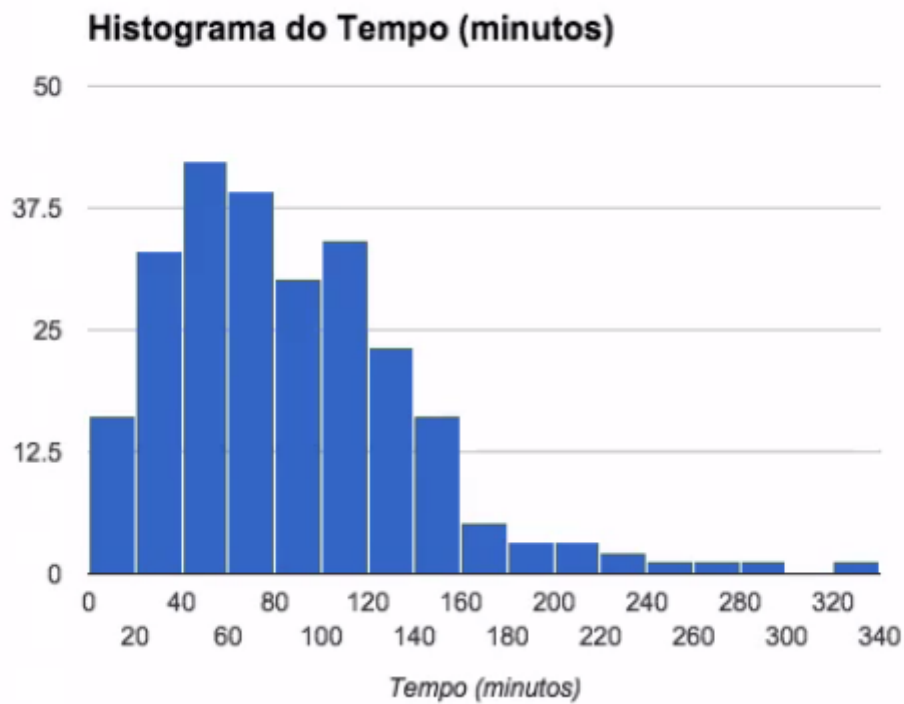
Distancia (km)	Tempo (minutos)
29	78
27	32
105	120
70	144
120	116
85	119
38	43
26	36
24	31
21	38
33	39
32	36
52	79
36	56
53	56
18	20
13	15
68	98
77	71

A planilha possui a distância em quilômetros e o tempo em minutos da entrega. Durante a semana, eles coletaram informações 250 eventos.

Primeiramente, isto não é uma série temporal. Não estamos analisando por datas, os valores estão soltos. A maneira como vamos tratar os dados será diferente da maneira como trataríamos uma série temporal. Poderíamos tratá-los como uma série temporal, se tivéssemos disponível o dia e a hora? Sim. Mas seriam informações irrelevantes para nós.

Iremos aqui trabalhar apenas como um conjunto de valores. Geralmente, nestes casos a primeira coisa que observamos é a distribuição. Ou seja, como os valores estão distribuídos, quais são os mais e menos comuns. Existe um gráfico chamado **histograma**, que irá desenhar justamente isto. Criaremos um para Distância e outro para o Tempo .





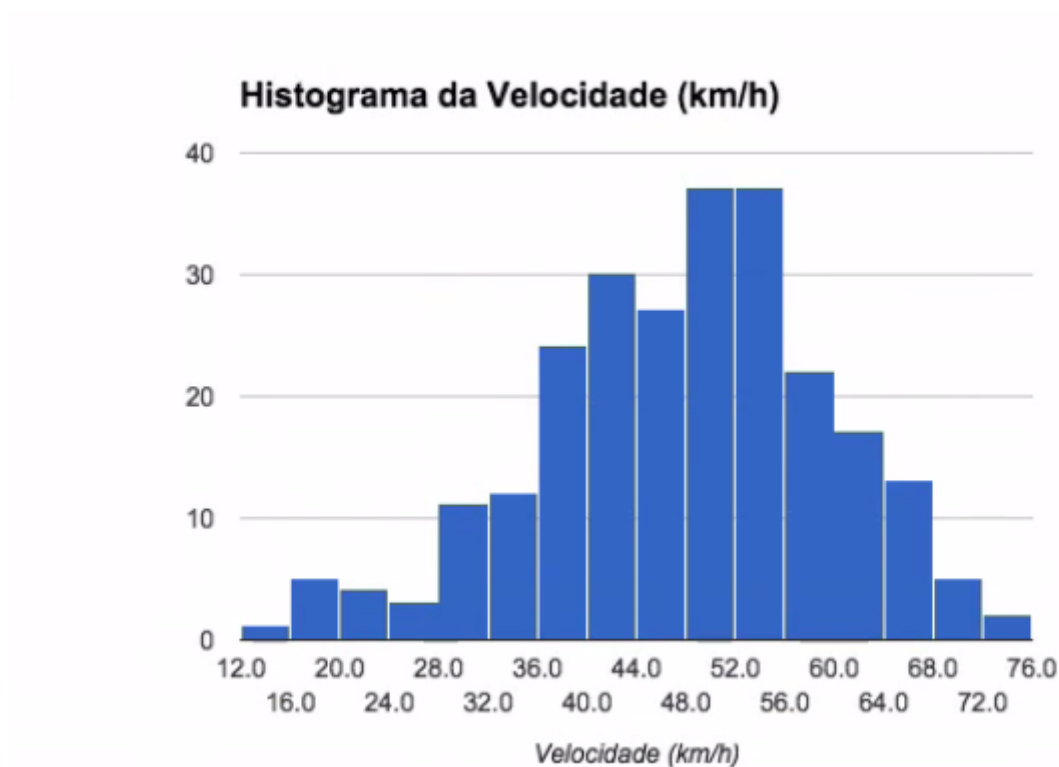
O eixo y sempre será referente a contagem.

Observe que as barras da distância são uniformes. No entanto, o histograma do tempo tem uma maior concentração até o tempo de 160 minutos e depois uma "cauda". Como analista nós temos que investigar o motivo de termos entregas que demoram até 340 minutos. Os motivos podem ser variados: a moto quebrou, o trânsito estava intenso ou tinha uma greve nas vias. É muito provável que algo aconteceu nestes eventos. Será preciso a coleta de novos dados para responder esta questão.

Nós também poderíamos calcular a velocidade média das motos durante a entrega. Encontramos estes valores dividindo a `Distância` pelo `Tempo`. Nossa sugestão é que você converta o tempo de **minuto** para **hora**.

Velocidade (km/h)
22.2
50.0
52.5
29.3
62.1
42.9
53.4
43.1
47.1
33.5
50.6
53.6
39.3
38.4
56.8
53.2
52.4
41.7
65.0

Criaremos também o histograma da Velocidade .



Agora, vemos uma cauda no lado esquerdo do histograma. Isto significa que alguns entregadores foram a uma velocidade inferior a 30km/h. Provavelmer

existiu algum problema, porque não é normal um entregador ir a 12km/h, por exemplo. Analisando também o gráfico da distância, vemos que tiveram entregas feitas em 120 km. Se a entrega foi feita em uma estrada, é normal que a velocidade seja superior a 60km/h.

As primeiras análises que podemos fazer é: a distância é constante, a maior parte das entregas são feitas em até 160 minutos, o que equivale a duas horas e meia. Vemos também que algumas entregas demoraram muito para serem feitas. Também percebemos que algumas entregas tiveram uma velocidade média muito baixa, enquanto a maior parte das entregas foram feitas com velocidades entre 40 e 68km/h.