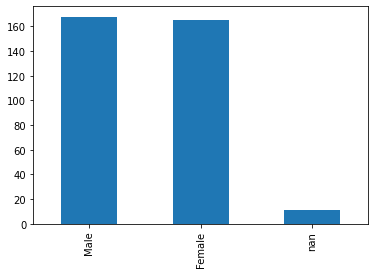
# Assignment 101

Als ik terugkijk op mijn uitwerking van wat Data Science precies is heb ik veel gebruik gemaakt van plaatjes. En een belangrijke quote: “Data 🡪 Knowledge”. Daarom zijn mijn PowerPoint dia’s heel abstract en bevatten ze veel informatie als er een goede presentator het woord heeft.

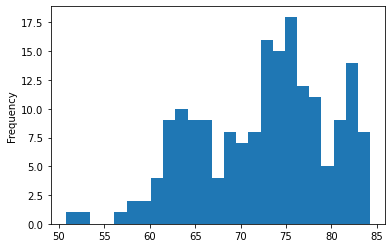
Met “Data 🡪 Knowledge” bedoelde we toen de tijd dat we met data verschillende dingen konden doen om daar zo mee knowledge te krijgen denk maar aan het voorspellen van data of data analyseren. In mijn mening is mijn visie niet veranderd en dus is er geen vernieuwde versie.

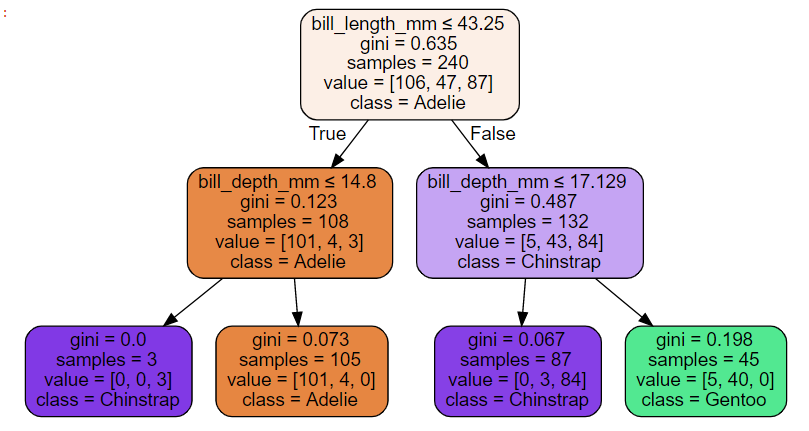
## Portofolio Assignments

Zoals eerder al is beschreven is een groot deel van Data Science het analyseren van gegevens. In mijn portofolio ben ik aan de slag gegaan met 2 grote datasets. De Pinguïn Dataset en een eigen gekozen Videogames Sales Dataset. Een analyse uitvoeren gaat niet zo makkelijk en er bestaan verschillende manier hoe en waarom je iets gaat vergelijken.

Ten eerste is er een “Univariate Analysis” gedaan. Op de “Categorical” Data kolommen. Daarmee kan je goed vergelijken wat het verschil is tussen de verschillende categorieën in je dataset. In de afbeelding hiernaast kun je goed de verhouding bekijken tussen male en female in de pinguïn dataset. Ook is er duidelijk te zien dat niet elke record in de set een geslacht heeft ingevuld.

Figuur

 Naast de “Categorical” analyse is er op de “Numerical” data ook een analyse gedaan. Met deze analyse kun je in plaats van categorische data, data verduidelijken met nummers erin. Er kan van alles uitgerekend worden van gemiddeldes, standaardafwijking, min, max, etc. al deze manieren van cijfers uitrekenen zegt op zichzelf niet zoveel daarom kan je met die data ook visualiseren zoals figuur 2. Kort om met een “Univariate Analysis” krijg je een duidelijk beeld over hoe je dataset in elkaar zit.

Een ander groot onderdeel van een Data Science analyse is het voorspellen van toekomstige gegevens. Met een ander term wordt het ook wel “Multivariate analysis” genoemd. Hierbij kan er een voorspelling gedaan worden als je data toevoegt wat de kans is dat het bij een bestaande categorie komt. Of als het gaat over numerical data wat het interval ervan is. Voorspellingen doen kan op vele manieren de manier die in het portfolio gebruikt wordt is een decision tree. Dat is eigen lijk een hoop if-statements achter elkaar gezet zodat je simpel gezegd kan bepalen waar de nieuwe data terecht komt. In figuur 3 is er een voorbeeld te zien van een decision tree.

Figuur

Figuur

# Assignment 102

In de vorige opdracht is al besproken wat Data Science is en al een aantal voorbeelden gegeven van het analyseren op een dataset. Een Data Scientist is eigenlijk degene die het uitvoert. Dus degene maakt analyses, voorspellingen etc. hij is dus instaat om grote hoeveelheden data om te zetten in bruikbare informatie. Een voorbeeld kan hiervan zijn dat degene kan voorspellen wat voor soort pinguïn er op een bepaald eiland gaat komen. Als data scientist kan je dus van alles voorspellen. In mijn portfolio is er bijvoorbeeld voorspeld wat het gewicht is van een bepaalde soort pinguïn. Of welke categorie game een Publisher in de toekomst mogelijk kan uitbrengen.