UPS BR Data Analyst Case

At UPS, Data Analysts are constantly challenged to provide insights and value to the company and our customers through open scope projects. The purpose of this case is to simulate that.

The case has three steps:

1. **Power Bi Challenge:** Develop a dashboard, connect it to the **online** dataset. provide visuals to meet the minimum requirements;

2. **Process Flow Challenge:** Create a flowchart based on a textual process description;

3. **Python Challenge:** Generate insights and visuals based on the provided dataset. **Plus:** Develop predictive model (classification) to identify if a shipment will arrive on time.

Deliverables:

- A detailed and well-organized **Power Bi file** (advice: use good practices on field names, formulas, etc.);

- A detailed process **Flowchart** in the format: PDF (PPT, Bizagi File, or any similar tool);

- A detailed and well-organized **Python Notebook** (.ipynb extension or Google Colab/Jupyter notebook link)

Below you will find the case description and more details about what we expect as your solution on each step. **Please, read carefully until the last page.**

Once completed, you may contact us via e-mail through the same address we have contacted you during the selection.

If you have any questions, please do not hesitate to contact us.

Be creative & have fun!

1. Power Bi Challenge

In this challenge, you are presented a sample dataset, that mocks fictitious parts shipping data. It is your challenge to understand the data and meet the minimum requirements. Notice that any insight/ideas beyond the request is a plus.

The database are the three following **csv** files:

- **shipments-details:** Fact table with shipment details;
- **parts-details:** Dimension table with shipped parts details;
- **fiscal-calendar:** Dimension table with the company fiscal calendar;

The **minimum** request for the dashboard is to help us to visualize the following points:

OVERALL

• Qty Shipped per Commodity (top 10);

• Value Shipped per Commodity (top 10);

• Country Shipment rank (Qty and Value);

• Average lead-time per country;

• City top offender lead-time;

• Number of orders;

• Average parts per shipment per country;

TREND

• Filters: Commodity, Country, City;

• Qty Shipped per Fiscal Quarter;

• Value Shipped per Fiscal Quarter;

• Average Lead Time per Quarter;

Database Source Setup

For the datasets "**parts-details**" and "**fiscal-calendar**", it is mandatory to use the **online file** as data source in the Power Bi database. Due to the file size, only the dataset "**shipments-details**" is allowed to use the connection in your local pc folder.

DATASET DETAILS

Shipments Details	
Field	**Description**
CUSTOMER_ORDER_NUMBER	Shipment identifier, customer call (order number)
DST_CITY	Destination city (city to which the part was shipped)
COUNTRY	Countrty where the part was shipped
PART_ID	Shippet Part ID (Part Identifier)
QTY_SHIPPED	Number of parts shipped
SHIPPED_DT	Date and time when the part was shipped
INVENTORY_YRL_COST	Part Inventory annual average cost
SHIPMENT_COST	Shipment average cost
POD_DT	Proof of delivery date and time (when the part is delivered)

Parts Details	
Field	**Description**
PART_ID	Part Identifier
COMMODITY	Part commodity classificatoin
PART_DESC	Part description
PART_VALUE	Part value in USD
PRODUCT_ID	Part commercial ID
PRODUCT_NAME	Part commercial name
SUPPLY_CHAIN_TYPE	Part supply chain category

Fiscal Calendar	
Field	**Description**
DATE	Calendar Date
FISCAL_YEAR	Company fiscal year
FISCAL_QUARTER	Company fiscal quarter
FISCAL_WEEK	Company fical week
FY_FQ	Concatenate fiscal year and fiscal quarter
FY_FW	Concatenate fiscal year and fiscal week

2. Process Flow Challenge

In order to improve customer experience and standardize shipping and packaging based on the company's global guidelines for unitization and repackaging, an audit process on the outbound line needs to be implemented.

It is your challenge to understand the audit process and **design a flowchart** to support the development of a tool (web application), that will be used by the operators on the outbound line. The tool will perform QA through questions about the package and store the answers in a database.

The Scope:

The operator checks Outbound Audit packaging aspects and marks in the tool if any of the below items failed:

- Appropriate labelling (UPS inbound label, UPS shipping label, RP/NRP information, required paper document (when applicable, e.g. battery) / sealing (box/envelope and part ESD/Bag void labels);
- Package integrity outside (complies with mechanical protection purposes, is whole, not damaged);
- Package visually adequate outside (by customer perspective), complying with recycling guidelines as per contract if box is reused;
- Package integrity inside (check internal part protection, when applicable to ensure no mechanical protection packaging parts are missing);
- WPB check (Part Number of the physical piece is the same as the Part Number from the package label and the same as the requested Part Number in the customer order);
- Scans the Box Code and the tool will confirm if the part audited is in the correct box as agreed in the “Box List per PN”;
- The tool shall need to check as first reference column named Global Box Part Number in “Box List per PN”. If Box Code is not found in that column, the tool shall look for that box part numbers by searching in other region’s columns;
- If actual Box Code is different from “Box List per PN” reference, tool will log a “fail” for this check and inform correct Box Code for repackaging action prior to release the dispatch;
- If actual Box Code is different from “Box List per PN” reference but Coordinator confirms that “Box List per PN” is inaccurate/wrong, auditor will flag “Box List Error” in the tool and inform actual Box Code used. The tool, in this case, will log a “pass” for this check and inform correct Box PN for repackaging action prior to release the dispatch;
- If no Box Code returns from “Box List per PN”, auditor will ensure Packaging Aspects are okay and record in the tool the Box Code used to dispatch this PN. The tool will log this case as “pass”;
- If actual Box Code matches “Box List per PN”, tool will log “pass” for the dispatch with same Box Code info recorded;
- If the Box Code is not outside of the box UPS will mentioned in tool no Box Code. If one of the items failed, tool will log “fail” for the dispatch.

GLOSSARY

Acronym	Description
PN	Part Number (Part Identifier)
RP	Returnable Part
NRP	Non-Returnable Part
WPB	Wrong Part Inbox
Box Code	Box identifier
Box Lisr per PN	List of proper packing box code by part number

BOX LIST PER PN EXAMPLE

PART NUMBER	GLOBAL BOX	LATAM BOX	EUROPE BOX	NORTH AMERICA BOX
14925	02087	PPL3P	0PRQ27	F22Q5
14954	AH273	PPL3P	0PRQ27	F22Q5
15203	AH273	PPL3P	0PRQ27	F22Q5
16635	AH273	PPL3P	Envelope	F22Q5
17466	02087	Envelope	Envelope	KHL19
18617	YPX32	Envelope	Envelope	KHL19
22918	YPX32	Envelope	3HXT44	KHL19
27792	YPX32	HGL3Q	3HXT44	YPVS18

3. Python Challenge

In this challenge, you are presented another fictitious data sample, that mocks customer purchases and deliveries. I this hypothetical scenario, the company wants to develop a machine-learning model to predict a delivery delay and notify the customer in advance.

Based on the file provided (predict-model-dataset.csv), the **minimum** request for this challenge is to answer the following questions using Python:

- Which Warehouse has more delays? Absolute and relative (%)
- Which Transportation mode has more delays? Absolute and relative (%)
- Generate a visuals to illustrate the distributions of "Gender" x "Customer_Rating"
- Based on the data provided, does package weight seems to be related to the delivery delay?

The **extra** request for this challenge is:

- Provide an overall data exploration, using **Python**. Metrics such mean, max, min, std_deviation;
- Predict model and its respective scores (Accuracy, Precision, Recall, F1-Score, ROC, Confusion Matrix, etc).

DATASET DETAILS

Predict Model Dataset	
Field	**Description**
ID	Customer Identifier
WAREHOUSE_BLOCK	Company warehouse block
MODE_OF_SHIPMENT	Transportation mode
CUSTOMER_CARE_CALLS	The number of calls made from enquiry for enquiry of the shipment
CUSTOMER_RATING	The company has rated from every customer. 1 (Worst) to 5 (Best)
COST_OF_THE_PRODUCT	Cost of the product in USD
PRIOR_PURCHASES	Number of prior Purchase
PRODUCT_IMPORTANCE	Company product importance category
GENDER	Customer gender
DISCOUNT_OFFERED	Discount offered on that specific product
WEIGHT_IN_GMS	It is the weight in grams
DELIVERY_DELAYED	Target variable, 1 if the product was NOT delivered on time, otherwise 0