

python データ加工研修 ガイダンス資料 (2020/6/25)イノベーション事業本部

三菱総研DCS株式会社



アジェンダ



進め方と事務連絡



全体説明





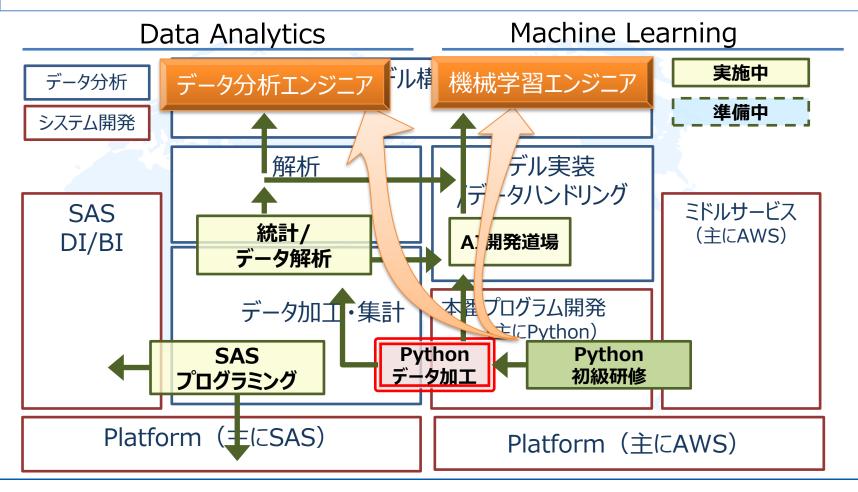
Pythonはデータ分析や機械学習の分野で広く使われているプログラム言語です。同分野でメジャーな言語の中でも汎用性が高い言語です。

- □ Pythonは科学技術計算に強い汎用プログラム言語です
 - 初版 (v0.9) は1991年に登場 (この頃は普通のプログラム言語)
 - 2000年代中盤頃からNumpy (行列計算)、pandas (データ解析)、scikit-learn (機械学習) などのライブラリが充実、一気に科学技術計算のスタンダードになりました
 - ◆ 統計言語であるSやRと異なり、あくまで汎用言語です
 - ◆ Django、FlaskなどのWebアプリケーションフレームワークもよく使われます
- 本研修でPythonを学び、データ分析や機械学習のプログラム開発のスタートラインに立ちましょう!
- □ ちなみに「Python」の名前の由来はイギリスのコメディアン Monty Pythonに由来
 - 迷惑メールの「spam」の由来もMonty Python



この研修で目指すゴール

この研修ではPythonでデータを扱う際に必要となるデータ加工処理の基礎スキルを習得します。Python初級研修と同様に基礎を身に着けるカリキュラムです。

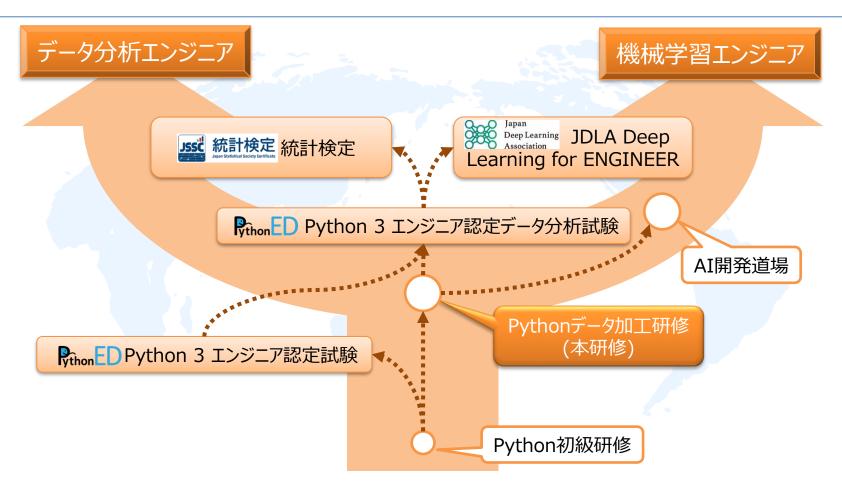




本研修と関連するベンダ資格

Pythonおよびデータ分析/機械学習に関連するベンダ資格も登場しています。

※ 本研修ではベンダ資格の取得を特に勧めてはいませんので、自己判断でご検討ください





進め方と事務連絡



学習スケジュール

Pythonデータ加工研修は、「テキスト学習」、「総合演習」、「修了判定」の3つのパートにわかれています。



テキスト学習

スライド資料(PDF)で文法等の解説、notebookで例題・演習を通して実際にコードを実行しながら研修を進めます。

● 総合演習

notebook上でより実践的な課題を解きます。

• 修了判定

テキスト学習・総合演習で学んだ内容を「知識テスト」と「コーディングテスト」で確認します。



タイムスケジュール(予定)

DAY1~DAY2は目安の時間です。DAY3総合演習およびDAY4修了判定テストは、時刻開始となります。お昼休みや休憩等については、各自で取得してください。

	DAY1 6月25日		DAY2 6月26日		DAY3 6月29日		DAY4 6月30日	
	カリキュラム	受講者	カリキュラム	受講者	カリキュラム	受講者	カリキュラム	受講者
09:00	準備	出欠応答	準備	出欠応答	準備	出欠応答	準備	出欠応答
09:30	Python基礎知識		データの加工		総合演習0		復習	
10:00	(2.5H)		(4.0H)		(1.0H)		(1.5H)	
10:30					総合演習0振返り			
11:00					総合演習1		終了判定テスト1	
11:30					(2.0H)		(1.0H)	回答提出 mail
12:00	昼休み		昼休み		昼休み		昼休み	
12:30	※休憩時間は適宜		※休憩時間は適宜		※休憩時間は適宜			
13:00	DataFrame(DF)		データ加工続き				終了判定テスト2	
13:30	の作成(1.0H)						(4.0H)	
14:00	DataFrame(DF)				総合演習1振返り			
14:30	の確認		データの要約・集計		総合演習2			
15:00	(3.0H)		(1.5H)		(1.5H)			
15:30								
16:00			データの可視化		総合演習2振返り			
16:30			(1.0H)		総合演習3Extra			回答提出 mail
17:00					復習		復習&振り返り	
17:30								



配布コンテンツ構成

配布したzipファイルをローカルのユーザーディレクトリに展開し、研修を開始してください。 主なフォルダ構成は下記の通りです。

Pythonデータ加工コンテンツ一式

- ├ 01_学習コンテンツ
 - ⊢ data
 - ├ Pythonデータ加工テキスト.pdf
 - ├ Python環境構築_00環境設定.pdf
 - └ 01~06までのJpyternotebookファイル
- ├ 02_総合演習.zip 🖰
- 03_修了判定テスト
 - ├修了判定テスト1.zip≜
 - └修了判定テスト2.zip傦
- └ 04 解答
 - ├ 01_テキスト演習問題解答
 - 02_総合演習解答
 - └ 総合演習0~2のzipファイル<mark>舎</mark>
 - 03_修了判定テスト解答.zip冎

※赤字骨となっているzipファイルは進行に応じてpassを連絡します。



学習の進め方:全体

基本的な進め方は自習です。標準PCにPython実行環境を用意して試しながら進めてください。

- □ 各日、研修開始時刻は9:30となります
 - Teams上で出欠確認を行います。9:30までに出欠確認の応答をお願いします。
 - 終了時間の目安は17:30となります。
- □ お昼休みおよび休憩
 - DAY 1 DAY2 のテキスト学習は、タイムスケジュールは目安です。お昼休み、休憩は各自で自由に取得してください。
 - DAY3 の総合演習は、タイムスケジュールに沿って解答が連携されます。
 - DAY4 の修了判定テストは、時刻開始となります。開始時間にはPC前で待機おねがいします。
- 困ったことがあったら
 - 疑問点、不明点については、Teams上に投稿してください。
 - 時間に余裕がある方は、他の受講者の質問に回答する等、助け合いの場として活用ください。
 - 研修スタッフも投稿を確認して、適宜回答いたします。
 - 研修の中で困ったことや相談事項がありましたら、研修スタッフへコメントください。



学習の進め方:テキスト学習

テキストを読み進めつつ、例題・演習をJupyter notebook上で解き、データ加工に必要な知識を身に着けていきます。

- □ テキストと各章のnotebookファイルで構成されています
 - テキストに沿って進めながら、対応するnotebookで例題・演習を解いてください。
 - データを誤って編集・保存してしまった場合は「01_学習コンテンツ > data >初期 data.zip」から初期のデータに戻すことができます。
- □ テキストやWebサイトを活用して進めてください
 - 本研修テキストも全て網羅しているわけではありません。不明点や疑問点等は積極的にWebサイトを活用して解消するようにしましょう。
- □ 演習回答は進度に応じて各自で確認してください
 - 解答は「04_解答 > 01_テキスト演習問題解答」配下に格納されています。



学習の進め方:総合演習

総合演習は、テキストで学んだ内容について理解を深める為のコンテンツです。

- □ 総合演習は4つのコンテンツで構成されています
 - パスワード付きのzipファイル(「02_総合演習.zip」)で圧縮されています。パスワードは総合演習開始時に、Teams上でご連絡しますので、展開して進めてください。
 - 使用するデータは上記zipファイルの中に梱包しています。
 - ◆ 総合演習0:処理の挙動の理解
 - ◆ 総合演習1:データ加工
 - ◆ 総合演習2:データ可視化
 - ◆ 総合演習3_Extra:加工したデータで予測モデル作成体験
- テキストやWebサイトを活用して進めてください
 - 総合演習実施時にテキストを使用しても構いません。
 - 本研修テキストも全て網羅しているわけではありません。不明点や疑問点等は積極的にWebサイトを 活用して解消するようにしましょう。
- 当日タイムスケジュールに従って解答を連携します。
 - 解答は「04 解答 > 02 総合演習解答」配下に演習毎にパスワード付きzipファイルで圧縮されています。各演習の終了時間に解除パスワードをTeams上で連携します。各演習後に振り返りの時間が設けられていますので、解答ファイルを確認しながら振り返りを進めてください。
 - ◆ 総合演習3_Extraについては、実行のみとなりますので解答はありません。



学習の進め方:修了判定テスト

修了判定テストは2部構成です。

修了判定テスト1(知識テスト:1h)と修了判定テスト1(コーディングテスト:4h)です。

- □ テキスト、Webの利用は可能です。
- 修了判定テスト1(知識テスト:1h)
 - 問題は「03_修了判定テスト >修了判定テスト1.zip 」にエクセルファイルとして格納されています。
 - 修了判定テスト1開始時間になりましたら、解除パスワードをTeamsで連携します。
 - テスト終了時間までに指定のメールアドレスに回答済みのExcelファイルを送付してください。
 - ◆ 時間内に終わらない場合でもその時点で一旦提出してください。
 - ◆ メールアドレスは別途Teamsで連絡します。
- □ 修了判定テスト2(コーディングテスト:4h)
 - 問題は「03_修了判定テスト >修了判定テスト2.zip」に格納されています。
 - 修了判定テスト2開始時間になりましたら、解除パスワードをTeamsで連携します。
 - テスト終了時間までに指定のメールアドレスに回答済みのnotebookファイルとCSVファイルを送付してください。
 - ◆ 時間内に終わらない場合でもその時点で一旦提出してください。
 - ◆ メールアドレスは別途Teamsで連絡します。



その他事務連絡

- 本研修に要した作業時間 (自習時間含む) は次のプロジェクトコードに賦課してください
 - プロジェクトコード:
 - ◆ OT-200 (【人財育成】各部人財育成(社外、部内))
 - 枝番:
 - ◆ 各部店指定の枝番 (部店ご担当者へご相談ください) ※ 指定が無ければ000としてください
 - プロジェクト作業名称:
 - ◆ Pythonデータ加工
- □ 労務管理の関係から、執務室またはテレワークで学習を進めてください
 - テレワークの方はテレワークの諸規定の遵守をお願い致します。
 - 常駐先の制約等でどうしても自宅で学習せざるを得ない場合はテレワークに準じる 扱いとなりますので、所属上長と予め相談しておいてください



それでは研修を開始してください。