# Programming for Data Science

# A Influência do Campeão Nacional de Futebol no Produto Interno Bruto de Portugal

David Oliveira[1], Filipe Marques[2], Henrique Pereira[3] & Manuel Oom[4]

[1]M20181430: M20181430@novaims.unl.pt

[2]M20181391: M20181391@novaims.unl.pt

[3]M20181395: M20181395@novaims.unl.pt

[4]M20181431: M20181431@novaims.unl.pt

**Abstract/Resumo:** Este estudo pretende demostrar qual a influência do vencedor da Liga Portuguesa de Futebol (LPF) na variação do Produto Interno Bruto (PIB). Este estudo nasce pela atenção dada em Portugal a este desporto e à potencial influência nos adeptos e equipas como consumidores e produtores de riqueza. Obtivemos vários dados de diversas fontes relativas à performance desportiva das equipas na LPF, e de dados macroeconómicos e sociais. Estes dados foram seleccionados e transformados a fim de obter as variáveis mais significantes para o estudo deste fenómeno. No decorrer do estudo, foi utilizada uma abordagem de modelação com regressão linear múltipla, pelo método Ordinary Least Squares (OLS). Concluímos que, apesar dos resultados mostrarem algumas correlações interessantes, não foi possível demonstrar que a equipa de futebol vencedora do campeonato nacional tem influência na variação do PIB de Portugal.

**Palavras-chave**: Liga Portuguesa de Futebol; Consumo Privado; Produto Interno Bruto.
**Declaração de contribuição**: Todos os intervenientes neste estudo tiveram igual contribuição.

## I. Introdução

O futebol tem uma presença relevante na população portuguesa. Tendo em conta o foco dado ao futebol no nosso dia-a-dia: em notícias espalhadas pelas diversas plataformas digitais; nos programas desportivos que ocupam uma grande parte da programação dos canais de televisão portuguesas; nas conversas entre amigos e colegas de trabalho, pode-se afirmar que se trata do desporto rei em Portugal. Este foco não é uma regra para toda a população, no entanto, todos acabam por ser afetados indiretamente pelo mediatismo deste desporto.
Em março de 2018, o Jornal de Notícias dava conta de que "*o futebol profissional contribui com 456 milhões para o PIB*" na época de 2016/2017, representando cerca de 0,25% do PIB nacional. Já em 2013, António Mexia (Presidente do Conselho de Administração da EDP) dizia que *"era bom para Portugal que o Benfica fosse campeão. Acho que isso tinha um efeito positivo no PIB*".
O PIB é um indicador macroeconômico que demonstra a criação de riqueza de um país. Este indicador representa o valor de bens e serviços produzidos no ano. Vários setores da sociedade e diversos fatores externos contribuem direta ou indiretamente para o PIB, razão pela qual poderá não ser possível quantificar a influência isolada do Futebol. Assim, para além do futebol, foram considerados outros factores como a inflação, os rendimentos e taxas de desemprego da população portuguesa.
O objectivo deste projecto é estudar a influência que a equipa de futebol vencedora do campeonato nacional tem no Produto Interno Bruto (PIB) de Portugal desse ano. Esta influência, a existir, poderá estar relacionada com uma maior propensão da população adepta a aumentar o consumo devido ao entusiasmo das vitórias dos seus clubes.

## II. Dados

Nesta secção será feita uma breve descrição dos dados utilizados, bem como da abordagem seguida. A abordagem foi composta pelas seguintes 3 principais fases:

- Extração e transformação de dados;
- Análise e seleção de variáveis; e
- Modelação.

# Programming for Data Science

### a. Extração e transformação de dados

Neste estudo foram utilizados dados com origem em diferentes fontes:

- www.zerozero.pt / www.transfermarkt.pt – Dados referentes aos clubes nacionais e dados do campeonato nacional de Futebol (vencedores, espectadores, entre outros);
- pt.investing.com/indices/ - dados relativos ao índice PSI 20; e
- www.pordata.pt – restantes dados, relativos ao PIB e à população portuguesa.

Na Tabela 1 é apresentada a natureza da informação utilizada para o estudo da influência dos vencedores do campeonato nacional de Futebol na evolução do PIB.

*Tabela 1 - Descrição dos dados utilizados no estudo*

| Natureza da informação | Descrição |
|---|---|
| Palmarés Liga Portuguesa (https://www.zerozero.pt/competicao_vencedores.php?id_comp=3) | Vencedores da liga Portuguesa de futebol em cada época, apresentando ainda o número de títulos de cada um dos vencedores, desde a época 1935/1936 até à actualidade |
| Espectadores Liga Portuguesa (https://www.transfermarkt.pt/liga-nos/besucherzahlenentwicklung/wettbewerb/PO1/plus/1) | Dados relativos aos espectadores nos estádios de futebol para cada época (a partir da época 2002/2003) |
| Histórico Classificação Liga Portuguesa (https://www.transfermarkt.pt/primeira-liga/ewigeTabelle/wettbewerb/PO1) | Histórico de todos os clubes participantes no campeonato nacional, com detalhes dos jogos (vitórias, empates, derrotas e diferença de golos) e respectivos pontos |
| PIB *per capita* (pib_per_capita.xlsx) | PIB *per capita* ao longo dos anos, desde 1960 até ao ano 2017 |
| Taxa de crescimento do PIB (taxa_crescimento_pib.xlsx) | Taxa de crescimento real do PIB ao longo dos anos, desde o ano 1961 até ao ano 2018 |
| População por grupo etário (populacao_grupo_etario.xlsx) | Informação da população residente em Portugal total, por grupo etário, desde 1970 até 2017 |
| População por género (populacao_genero.xlsx) | Informação relativa ao número de residentes em Portugal, por género |
| Euribor (Taxas-de-juro-Euribor.xlsx) | Valores das taxas de juro indexantes (Euribor) desde 1999 até 2018 |
| PSI20 (PSI20.xlsx) | Informação histórica do principal índice de referência do mercado de capitais português (PSI20) desde 1993 até ao presente |
| Consumo Privado em % do PIB (consumoprivado_pib.xlsx) | Dados históricos do consumo privado em percentagem do PIB, identificando ao longo dos anos (1960-2018) a percentagem do PIB que corresponde a consumo privado |
| Inflação (inflacao.xlsx) | Taxa de Inflação (Taxa de Variação do Índice de Preços no Consumidor) em valores totais e por consumo individual por objectivo, desde 1978 até 2018 |
| Desemprego (desemprego_percentagem.xlsx) | Informação relativa à percentagem de desemprego anual em Portugal, total e por género, desde 1983 até 2018 |
| Ganho dos trabalhadores (ganho_por_sexo.xlsx) | Dados do ganho médio mensal dos trabalhadores por conta de outrem, sendo apresentada a informação pela população total e por género, desde 1985 até 2017 |
| Remuneração por sexo (remuneracao_por_sexo.xlsx) | Informação referente à remuneração média mensal dos trabalhadores por conta de outrem, população total e por género, desde 1985 até 2017 |
| Rendimento em % do PIB (rendimento_empercentagem_pib.xlsx) | Dados do rendimento nacional bruto e rendimento disponível bruto em % do PIB ao longo dos anos, desde 1960 até ao ano 2017 |

Foram usadas diferentes formas de extração e leitura de dados, nomeadamente: leitura directa de dados de ficheiros Excel, através do *package* Pandas, obtidos a partir de diversas fontes de informação; extracção de informação a partir da web, utilizando a *package* BeautifulSoap – este *package* permite a leitura da informação de url, possibilitando a eliminação do processo de extracção manual dos dados das fontes de informação.

De forma a garantir a coerência da informação em análise, foram aplicadas transformações aos dados recolhidos, nomeadamente:

- **Alteração do tipo de dados** – alteração de tipos de dados desadequados à natureza da informação subjacente (e.g. substituição de "*objects*" por "*float64*" incluindo a eliminação de caracteres '.' ou aplicando a substituição de caracteres ',' por caracteres '.'; substituição de "*objects*" por "*datetime*";
- **Criação de variáveis –** A variável "época desportiva" é composta pelos dois últimos algarismos do ano de início e de fim da época, separados por '/' (e.g. 04/05 corresponde à época de 2004/2005). Para além da transformação natural desta variável, uma vez que nas épocas mais recentes o processo de importação considerou como sendo uma data no formato mês/ano, foram separados os anos de início e de fim da época em duas novas variáveis;
- **Renomear colunas –** algumas colunas são renomeadas de forma a manter uma linguagem mais clara na apresentação do *dataset* e a assegurar uma leitura mais simples dos dados;
- **Agrupar linhas –** algumas fontes de dados continham dados cuja informação útil deriva da agregação de linhas, utilizando a média, o máximo ou o mínimo (e.g. dados do psi20 foram obtidos por mês e necessitamos da granularidade por ano, de forma a mapear com a chave do *dataset*); e
- **Transformação de variáveis categóricas** - A equipa vencedora, como variável categórica, foi depois transformada em variáveis *dummies* para efeitos de modelação. Como temos 4 equipas vencedoras do campeonato iremos eliminar a *dummy* relativa ao Boavista para obtermos N-1 *dummies* a partir de N categorias, mas também pela frequência baixa de vitórias desta equipa.
- **Transformação de escala** – Algumas variáveis foram transformadas, de forma a tornar a escala das variáveis o mais similar possível. Para este feito foi aplicada a transformação logarítmica à variável total de população e à variável de variação do PSI foi aplicada uma transformação nas unidades, através da divisão por 10.

De forma a manter o *Jupyter Notebook* organizado, foi desenvolvida uma função de extração e transformação para cada uma das fontes de dados. Estas funções encontram-se implementadas na biblioteca *pre_processing_lib.py*.

### b. Análise e seleção de variáveis

As variáveis obtidas das várias fontes de informação foram agrupadas num único *dataframe*, chamado dataset, de forma a permitir uma análise conjunta de toda a informação recolhida. O dataset tem como *index* o ano, que corresponde ao ano de medição dos índices utilizados e ao ano de final de época das épocas desportivas (e.g. ano 2018 para a época 2017/2018). Considerou-se o ano de fim da época desportiva, por ser o ano em que se celebra a vitória do campeonato relativo à época.
O dataset original contém 85 casos, correspondendo a um período global de 1935 a 2019, com 50 variáveis. Da análise às variáveis, verificou-se que nenhuma contém dados para a totalidade desse período. Foi considerada como variável alvo a "*Tx_Cresc_Real_PIB*", correspondente à taxa real de variação do PIB por ano, face ao ano anterior. Esta variável tem dados para o período de 1961 a 2018, fixando assim o período máximo possível de análise.
Existem, contudo, variáveis que nos pareceram interessantes e que apenas têm dados para um período temporal mais reduzido (e.g. PSI20_Variacao – 1993 a 2019 ou Euribor 6 meses -1999 a 2018) o que obrigou à realização de análises com diferentes amplitudes temporais, com a utilização de menos variáveis em períodos mais longos.
Da análise preliminar às variáveis, percebemos ainda que algumas variáveis estão correlacionadas (e.g. população total vs população feminina ou masculina) ou não são relevantes para o este estudo (e.g. total de espectadores por equipa). Pelo princípio do *garbadge in garbadge out*, incluir as variáveis referidas atrás poderia colocar em causa o resultado do nosso modelo explicativo da variável alvo. Por esse motivo, eliminaram-se variáveis redundantes e/ou não pertinentes, ficando o dataset com um total de 12 variáveis.
Como resultado da fase de seleção de variáveis, resultaram 3 *subdatasets* de dados, um com variáveis com dados até 57 anos, outro com variáveis com dados até 25 anos e um último com variáveis com dados até19 anos:

*Tabela 2 - Subdatasets utilizados e respectivas variáveis*

| Subdatasets (amplitude de anos) | Variáveis seleccionadas |
|---|---|
| *Subdataset* com 57 anos (1961 to 2017) | EquipaVencedora; Tx_Cresc_Real_PIB; TotalPop_Log; RendDisponivelPCT_PIB; Consumo_Privado_em_Pct_PIB |

| | |
|---|---|
| *Subdataset* com 25 anos (1993 to 2017) | EquipaVencedora; Tx_Cresc_Real_PIB; TotalPop_Log; PSI20_Variacao10x; RendDisponivelPCT_PIB; Consumo_Privado_em_Pct_PIB; Inflacao_Total; Inflacao_Lazer_recreação_cultura; DesempregoPCT_Total |
| *Subdataset* com 19 anos (1999 to 2017) | EquipaVencedora; Tx_Cresc_Real_PIB; TotalPop_Log; Euribor 3 meses; Euribor 6 meses; Euribor 12 meses; PSI20_Variacao10x; RendDisponivelPCT_PIB; Consumo_Privado_em_Pct_PIB; Inflacao_Total; Inflacao_Lazer_recreação_cultura; DesempregoPCT_Total |

### c. Modelação

Após obtenção dos subdatasets de dados a estudar, foram implementados vários modelos de regressão linear múltipla, para tentar comprovar e explicar a influência que a equipa de futebol vencedora do campeonato nacional tem no Produto Interno Bruto (PIB) de Portugal desse ano.
Foi utilizado o package *StatsModels* para implementar os vários modelos lineares, tendo sido criados 6 modelos para cada subdataset. Os modelos de regressão linear múltipa foram implementados pelo método Ordinary Least Squares (OLS).
A análise do resultado dos modelos foi realizada tendo em conta:

i. Diagnóstico de cada modelo considerando os seguintes parâmetros em conjunto:
   a. O maior valor do $R^2$ ajustado (modelos lineares múltiplos);
   b. O menor número de parâmetros (β) de variáveis não significativos (IC a 95%);
   c. O maior valor de F e a sua significância (IC a 95%);
   d. O menor valor da Soma dos Quadrados dos Resíduos (SSR)
   e. O menor valor do 'Condition Number' (multicolinearidade reduzida)
   f. A estatística de Durbin-Watson entre valores 1.5 a 2.5 (baixa autocorrelação)
ii. Comparação entre o resultado do diagnóstico de cada modelo;
iii. Validação dos restantes requisitos afectos aos erros do modelo mais favorável:
   a. **L** (*Linearity*): Linearidade das variáveis explicativas face à variável alvo;
   b. **I** (*Independence*): Os erros devem ser independentes;
   c. **N** (*Normality*): Os erros entre devem ser normalmente distribuídos; e
   d. **E** (*Equality of variance*): A variância dos erros deve ser constante.

Os modelos foram construídos sequencialmente, de uma forma geral, começando pelo maior número de variáveis, com sucessivas eliminações/modelações de variáveis favoráveis face ao modelo anterior, até um limite de 6 modelos por cada subdataset.

## III. Resultados e Discussão

**Subdataset a 57 anos**
Analisando o *subdataset* com um período de 57 anos, verificamos que a nossa variável alvo tem correlação linear positiva com o rendimento disponível em percentagem do PIB (R=0,6), negativa com o total da população (R=-0,6). Adicionalmente, embora a variável consumo privado em percentagem do PIB não tenha uma correlação forte com a variável alvo (R=0,1), verificamos que esta correlação linear é notória se a análise for realizada por equipa vencedora. Neste caso, verificamos uma correlação positiva quando a equipa vencedora é o Sporting e negativa quando é o Benfica ou o Porto (vide *Jupyter Notebook* II.1.2.).

**Subdataset a 25 anos**
Analisando o *subdataset* com um período de 25 anos, verificamos que a nossa variável alvo tem correlação linear positiva com o rendimento disponível em percentagem do PIB (R=0,4) e, embora menos notória, com a variação do PSI20 (R=0,2). Existe também, uma correlação linear negativa com o total da população (R=-0,4), com o consumo privado em percentagem do PIB (R=-0,5) e com o desemprego (R=-0,6).

**Subdataset a 19 anos**
Analisando o *subdataset* com um período de 19 anos, verificamos que a nossa variável alvo tem correlação linear positiva com a taxa Euribor a 3, 6 e 12 meses (R=0,4), rendimento disponível em percentagem do PIB (R=0,4), negativa com o total da população (R=-0,5), consumo privado em percentagem do PIB (R=-0,4), desemprego (R=-0,6).

**Modelos**

Do subdataset a 57 anos podemos obter o seguinte resumo dos modelos analisados:

*Tabela 3 - Modelos analisados para o dataset a 57 anos*

| | ANOVA | | Modelo Linear | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Nº Modelo** | Graus L. (resid.) | SSR | $R^2$(ajust) | F | P (F) | Condition Number | Nº Param. p > 0.05 | Durbin-Watson |
| **0** | 54 | 601.42 | 0.484 | 18.87 | < 0.05 | 1.8 | 0 | 0.94 |
| **1** | 52 | 559.95 | 0.067 | 2.01 | 0.106 | 85.8 | 5 | 0.921 |
| **2** | 51 | 408.09 | 0.306 | 5.96 | < 0.05 | 128.9 | 5 | 1.12 |
| **3** | 51 | 403.23 | 0.315 | 6.15 | < 0.05 | 326.3 | 5 | 1.121 |
| **4** | 53 | 417.79 | 0.317 | 9.67 | < 0.05 | 308.5 | 2 | 1.198 |
| **5** | 55 | 450.33 | 0.290 | 23.95 | < 0.05 | 283.1 | 0 | 1.156 |

Verificamos que o modelo mais favorável é o modelo nº 5.

***Modelo 5:*** $Tx\_Cresc\_Real\_PIB = \beta_0 + \beta_1 \times TotalPop\_Log \times RendDisponivelPCT\_PIB + \varepsilon$

Examinando o diagnóstico dos resíduos (vide *Jupyter Notebook* III.1) verificamos algumas condições LINE, apesar de demonstrarem a existência de possíveis *outliers* nos anos 1962 e 1965 (extremos positivos) e 1975 (extremo negativo). Contudo também se constata que os resíduos não foram totalmente minimizados, e que existe, pelo valor de Durbin-Watson, autocorrelação positiva ligeira. Neste modelo não há significância do peso/impacto da variável respeitante à equipa vencedora na variável dependente.

Do subdataset a 25 anos podemos obter o seguinte resumo dos modelos analisados:

*Tabela 4 - Modelos analisados para o dataset a 25 anos*

| | ANOVA | | Modelo Linear | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Nº Modelo** | Graus L. (resid.) | SSR | $R^2$(ajust) | F | P (F) | Condition Number | Nº Param. p > 0.05 | Durbin-Watson |
| **0** | 14 | 45,00 | 0,354 | 2,32 | 0,07 | 29909,4 | 11 | 2,04 |
| **1** | 17 | 46,01 | 0,456 | 3,88 | < 0.05 | 9895,8 | 5 | 2,02 |
| **2** | 18 | 46,04 | 0,486 | 4,79 | < 0.05 | 9859,9 | 4 | 2,02 |
| **3** | 18 | 48,72 | 0,456 | 4,36 | < 0.05 | 8561,8 | 3 | 1,81 |
| **4** | 18 | 33,53 | 0,626 | 7,69 | < 0.05 | 1112800,1 | 1 | 2,10 |
| **5** | 23 | 82,16 | 0,282 | 10,44 | < 0.05 | 5,6 | 0 | 1,11 |

Verificamos que existem 2 modelos que parecem mais favoráveis relativamente aos restantes, o modelo nº 4 e nº 5, respectivamente:

***Modelo 4:*** $Tx\_Cresc\_Real\_PIB = \beta_0 + \beta_1 \times TotalPop\_Log + \beta_2 \times TotalPop\_Log \times Consumo\_Privado\_em\_Pct\_PIB + \beta_3 \times Inflacao\_Total + \beta_4 \times DesempregoPCT\_Total + \beta_5 \times EquipaVencedora\_Benfica + \varepsilon$

***Modelo 5:*** $Tx\_Cresc\_Real\_PIB = \beta0 + \beta1 \times TotalPop\_Log \times DesempregoPCT\_Total + \varepsilon$

Embora nenhum dos dois modelos cumpra na totalidade os requisitos LINE, o modelo nº 5 aparenta ser o melhor, se tivermos em conta o valor de F e a ausência de parâmetros não significativos e de multicolinearidade. Não obstante tem o pior SSR da série, o pior $R^2$ ajustado e uma ligeira autocorrelação positiva. Já o modelo nº 4, tem melhor comportamento relativamente aos resíduos, comparativamente ao modelo 5, uma vez que aparentam ter uma independência mais significativa.

Do subdataset a 19 anos podemos obter o seguinte resumo dos modelos analisados:

*Tabela 5 - Modelos analisados para o dataset a 19 anos*

| | ANOVA | | Modelo Linear | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Nº Modelo** | Graus L. (resid.) | SSR | $R^2$(ajust) | F | P (F) | Condition Number | Nº Param. p > 0.05 | Durbin-Watson |
| **0** | 6 | 16,44 | 0,408 | 2,04 | 0,197 | 3395,8 | 13 | 2,58 |
| **1** | 7 | 19,15 | 0,409 | 2,13 | 0,162 | 2071,3 | 12 | 2,29 |
| **2** | 9 | 20,32 | 0,513 | 3,10 | 0,053 | 1672,9 | 9 | 2,34 |
| **3** | 11 | 20,69 | 0,594 | 4,76 | < 0.05 | 1625,6 | 6 | 2,34 |
| **4** | 13 | 21,85 | 0,637 | 7,32 | < 0.05 | 771,3 | 4 | 2,55 |
| **5** | 15 | 25,72 | 0,630 | 11,21 | < 0.05 | 654,1 | 0 | 2,27 |

Verificamos que o modelo mais favorável é o modelo nº 5.

***Modelo 5:*** $Tx\_Cresc\_Real\_PIB = \beta_0 + \beta_1 \times TotalPop\_Log \times RendDisponivelPCT\_PIB + \beta_2 \times Euribor3meses + \beta_3 \times EquipaVencedora\_Benfica + \varepsilon$

Este modelo tem o segundo melhor $R^2$ ajustado, o melhor valor de F, a ausência de parâmetros das variáveis não significativos, a ausência de autocorrelação e a menor evidência de multicolinearidade. Não obstante, tem o pior SSR, e analisando os gráficos dos resíduos (vide *Jupyter Notebook* III.3.) não se verifica o cumprimento dos requisitos LINE.

## IV. Conclusões

Não foi possível demonstrar que a equipa de futebol vencedora do campeonato nacional tem influência na taxa de variação do Produto Interno Bruto de Portugal.

Foram desenvolvidos 6 modelos de regressão linear múltipla, para diferentes períodos de tempo e com recurso a diferentes variáveis, e nenhum dos modelos obtidos cumpre com os requisitos pressupostos para a utilização de modelos de regressão linear (LINE). Assim, não podemos excluir a nossa hipótese nula, de que a equipa vencedora não influencia a taxa de variação do PIB.

Embora existam notícias que dão conta de que o futebol profissional contribuí com 0,25% do PIB, o facto é que não foi possível verificar a influência que o vencedor do campeonato tem no PIB. Esta conclusão pode dever-se a:

i. A equipa de futebol vencedora do campeonato nacional não influencia, nem tem nenhuma relação, com o PIB de Portugal;
ii. Os 0,25% do valor do PIB influenciados pelo futebol nacional serem independentes do valor absoluto deste indicador, fazendo com que a previsão da variação desta pequena parcela não revele a previsão da variação global do PIB; ou
iii. A abordagem utilizada, modelo de regressão linear, não ser a adequada para analisar o fenómeno concreto em estudo.

Como sugestões de futuros trabalhos, poderá ser explorada a combinação das variáveis utilizadas com outras não incluídas neste estudo, ou mesmo a análise dos mesmos dados utilizados neste estudo numa série temporal. Poderá também ser analisado futuramente a influência que a equipa vencedora do campeonato tem na parcela do PIB relativa ao futebol.

## V. Referências


[1] https://www.jn.pt/desporto/interior/futebol-profissional-contribui-com-456-milhoes-para-o-pib-9210445.html, notícia publicada a 23 março 2018.

[2] https://www.publico.pt/2018/03/23/desporto/noticia/futebol-portugues-contribuiu-com-456-milhoes-de-euros-para-o-pib-1807859, notícia publicada a 23 março 2018.

[3] https://tribunaexpresso.pt/futebol-nacional/2019-04-03-Contributo-do-futebol-portugues-para-o-PIB-baixa-13-para-396-milhoes-na-ultima-epoca, notícia publicada a 3 abril 2019.

[4] https://www.publico.pt/2004/06/20/jornal/a-excessiva-importancia-social-e-politica-que-damos-ao-futebol-e-ditatorial-e-asfixiante-189923, artigo publicada a 20 junho 2004.

[5] Kutner, M.H., Nachtsheim, C.J., Neter, J., Li,W. (2004) Applied Linear Statistical Models, 5th edition, McGraw-Hill.

[6] McKinney, Wes. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. "O'Reilly Media, Inc.", 2012.